INTEGRATIVE BRAIN FUNCTION DOWN UNDER

EDITED BY: Greg Stuart, Pankaj Sah and Gary F. Egan

frontiers Research Topics

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

INTEGRATIVE BRAIN FUNCTION DOWN UNDER

Topic Editors:
Greg Stuart, Australian National University, Australia
Pankaj Sah, The University of Queensland, Australia
Gary F. Egan, Monash University, Australia

Table of Contents

frontiers
in Neural Circuits

Editorial: Integrative Brain Function Down Under

Pankaj Sah[1]*, Greg J. Stuart[2] and Gary F. Egan[3]

[1] *Queensland Brain Institute, The University of Queensland, Brisbane, QLD, Australia,* [2] *John Curtin School of Medical Research, Australian National University, Canberra, ACT, Australia,* [3] *Monash Biomedical Imaging and School of Psychological Sciences, Monash University, Melbourne, VIC, Australia*

Editorial on the Research Topic

Integrative Brain Function Down Under

Despite decades of research, how the brain interacts with the world is still one of the greatest challenges of the twenty-first century. To address this challenge, Australia's leading brain researchers are investigating brain structure and function at multiple scales, from single cells and synapses, to circuits and networks, to whole brain systems.

In this Research Topic, we are pleased to present a collection of articles and reviews from Australian neuroscientists, engineers, and computer scientists, covering a multitude of topics ranging from ion channel function in single neurons to sensory information processing within neural networks. The aim of this research is to increase our understanding of how the brain integrates information across multiple levels. Key to this process is the development of new experimental instruments and computational tools. Hopefully, the new insights obtained will also aid development of approaches to repair and restore function to the damaged brain.

In order to understand integrative brain function, it is critical to understand how the brain processes information. Information processing in the brain relies on spiking activity in single neurons, which requires the movement of charged ions through ion channels in neuronal membranes. Autuori et al. investigate ion flow through small conductance calcium-activated potassium channels (so-called SK channels), which contribute to the after hyperpolarization that follows spike activity in many neuronal cell types. The team identified that the rSK1 protein acts as a chaperone for rSK2 channels, indicating that expression of the rSK1 gene may control the level of functional SK current in neurons. To further gain insight into information processing in the brain, an understanding of how populations of neurons encode information in their patterns of spiking activity is essential. Triplett and Goodhill review recent methods for extracting variables that quantitatively describe how sensory information is encoded. In particular, they discuss methods for estimating receptive fields, modeling neural population dynamics and inferring low dimensional latent structure from neuronal populations. In a related study, Zavitz et al. present an overview of some of the most promising analytical approaches for making inferences from population recordings in multiple brain areas, such as dimensionality reduction and changes in correlated variability. Hadjidimitrakis et al. review the evidence related to functional communication between subregions of the posterior parietal cortex and how recordings in this region can be used to decode movement goals. These data suggest that the posterior parietal cortex works as a dynamic network of sensorimotor loci that combine multiple signals which work in concert to guide motor behavior, and raises the possibility of using parietal neuron activity to better drive neuroprosthetic devices for motor control.

The brain integrates and processes a massive amount of sensory information to guide behavior crucial for survival. Oestreich et al. investigate the connections between brain regions activated by speech. They identified that structurally connected brain regions are also functionally engaged by externally-generated and temporally-predictable speech patterns. The research provides evidence that the brain continually predicts incoming sensory events based on past experience in order to respond to unexpected events in a fast and efficient manner. Work by Lian et al. explores how the visual system codes visual stimuli. Using a biologically plausible model they show how complex experimental phenomena, such as the shape of receptive fields and contrast invariance of orientation tuning, can be implemented in primary visual cortex by sparse coding. In a related article, Chaplin et al. compare the representations of space and motion in the visual and auditory cortex, and examine how single neurons in these two areas encode the direction of motion. They discuss how humans integrate audio and visual motion cues, and the regions of the cortex that may mediate this process.

Several articles in this collection are dedicated to developing new approaches and building better models of integrative brain function. For example, Vidyasagar et al. propose a model of how the claustrum orchestrates and integrates activity across different cortical areas by boosting synchronized oscillations between these areas. Gollo et al. present a non-linear hierarchical model that provides unique insights into the brain architecture underlying the representation and appraisal of perceptual belief and precision in the prefrontal cortex. Jacques et al. describe a novel approach to precisely map molecular markers in neuronal networks through quantitative topographic measurement. This approach can be used to gain a greater understanding of functional encoding within sub-nuclei during memory formation and may prove advantageous for studying the cellular basis of addiction as well as pathological memory models. Finally, Arnatkevičiute et al. review studies investigating gene expression patterns associated with hub connectivity in neural networks and present evidence that some of these expression patterns are conserved across species and scales. Together, these studies provide new models of brain networks which aid our understanding of how the brain integrates information across multiple brain regions.

The articles in this Research Topic study the relationship between brain activity and behavior at multiple spatial and temporal scales—from single cell electrical and biochemical activity to patterns of activity in large scale circuits and networks. In doing so they help to build an integrated model of how the brain processes information and thereby contribute to a deeper understanding of how the brain interacts with the world.

AUTHOR CONTRIBUTIONS

PS initiated the Editorial draft, to which both GS and GE contributed and edited. All authors contributed to the article and approved the submitted version.

FUNDING

frontiers
in Neural Circuits

Prediction of Speech Sounds Is Facilitated by a Functional Fronto-Temporal Network

Lena K. L. Oestreich [1,2]*, Thomas J. Whitford [3] and Marta I. Garrido [1,2,4,5]

[1]Queensland Brain Institute, The University of Queensland, Brisbane, QLD, Australia, [2]Centre for Advanced Imaging, The University of Queensland, Brisbane, QLD, Australia, [3]School of Psychology, University of New South Wales, Sydney, NSW, Australia, [4]Australian Centre of Excellence for Integrative Brain Function, The University of Queensland, Brisbane, QLD, Australia, [5]School of Mathematics and Physics, The University of Queensland, Brisbane, QLD, Australia

Predictive coding postulates that the brain continually predicts forthcoming sensory events based on past experiences in order to process sensory information and respond to unexpected events in a fast and efficient manner. Predictive coding models in the context of overt speech are believed to operate along auditory white matter pathways such as the arcuate fasciculus and the frontal aslant. The aim of this study was to investigate whether brain regions that are structurally connected via these white matter pathways are also effectively engaged when listening to externally-generated, temporally-predicable speech sounds. Using Electroencephalography (EEG) and Dynamic Causal Modeling (DCM) we investigated network models that are structurally connected via the arcuate fasciculus from primary auditory cortex to Wernicke's and via Geschwind's territory to Broca's area. Connections between Broca's and supplementary motor area, which are structurally connected by the frontal aslant, were also included. The results revealed that bilateral areas interconnected by indirect and direct pathways of the arcuate fasciculus, in addition to regions interconnected by the frontal aslant best explain the EEG responses to speech that is externally-generated but temporally predictable. These findings indicate that structurally connected brain regions involved in the production and processing of auditory stimuli are also effectively connected.

Keywords: predictive coding, electroencephalography (EEG), dynamic causal modeling (DCM), effective connectivity, structural connectivity

INTRODUCTION

The ability to predict imminent sensations from past experiences such as hearing a familiar song, is crucial to efficiently process the abundance of sensory stimulation we experience at any moment. Moreover, it enables rapid detection of unexpected events and facilitates adaption to novel contingencies in our environment (Mumford, 1991, 1992). The predictive coding framework posits that in an effort to optimize sensory processing, the brain continuously generates models of the environment that are based on memories specific to a given context (Friston, 2005; Garrido et al., 2007). According to this theory, predictions are generated in higher cortical areas and communicated to lower sensory areas via backward (top-down) connections. The sensory areas then compare actual sensory input with the predicted sensation and its difference, i.e., mismatch or prediction error, is conveyed upstream via forward (bottom-up) connections (Rao and Ballard, 1999). This prediction error signal facilitates continuous updating of the internal predictive model.

The functional anatomy underlying auditory prediction is yet to be conclusively determined. One of the primary ways that humans produce sounds is by vocalizing (e.g., speaking). It is plausible that the neural architecture involved in producing and perceiving willed speech overlaps with the neural architecture involved in predicting sounds more generally (Gagnepain et al., 2012). The arcuate and aslant fasciculi are two white matter fiber bundles that are potentially involved in predictive coding in the context of willed speech. The arcuate fasciculus provides a direct connection between speech production (Broca's) and speech perception (Wernicke's) areas. In addition to direct, long segment fibers connecting Broca's and Wernicke's area, the arcuate fasciculus also has shorter, indirect connections consisting of an anterior pathway which connects Broca's area to Geschwind's territory, and a posterior pathway which connects Geschwind's territory and Wernicke's area (Catani et al., 2005). These long and short distance pathways of the arcuate fasciculus possess different functional roles: whereby the direct pathway is thought to be involved in phonological functions, the indirect pathways are associated with semantic functions (Catani and ffytche, 2005). Specifically, the posterior indirect pathway is thought to be involved in auditory comprehension and the anterior indirect pathway in the vocalization of semantic information (Catani et al., 2005). Evidence for a role of the arcuate fasciculus in predictive coding in the context of willed speech comes from studies with schizophrenia patients (Whitford et al., 2017), which showed that the structural integrity of the arcuate fasciculus is associated with predictive coding deficits, as quantified by the level of electrophysiological suppression to willed speech. The frontal aslant, which directly connects Broca's area with the supplementary motor area (SMA; Catani et al., 2012) may also play a role in predictive coding in the context of speech production, as it is known to be involved in verbal fluency (Catani et al., 2013) and speech initiation (Fujii et al., 2016).

According to the "forward model" of speech production, the sensory consequences of self-generated speech are predicted through a copy of the motor command, which is sent via top-down projections from the motor cortex to the sensory system (Houde and Jordan, 1998). If the mechanisms involved in predictive coding of external, predictable sounds operate via similar neural pathways as those involved in predictive coding of willed speech, then the former may rely on the functional engagement of the arcuate fasciculus and the frontal aslant.

In this study, we formulated a set of dynamic causal models (DCMs) to investigate the functional underpinnings of auditory prediction of external, predictable speech sounds. These DCMs included brain regions interconnected via the arcuate fasciculus and the frontal aslant. It was hypothesized that models with both forward (bottom-up) and backward (top-down) connections, which convey sensory input and prediction, respectively, would perform better than models with forward (bottom-up) connections alone. Furthermore, we explored whether auditory prediction was better explained by alternative models that included or excluded the above mentioned regions along the arcuate fasciculus (Geschwind's territory) and the frontal aslant (SMA).

MATERIALS AND METHODS

Participants

Seventy-five healthy participants (38% males, aged 18–44 years, 95% right-handed) were recruited through the online recruitment systems SONA-1 and SONA-P at the University of New South Wales (UNSW), Australia. Participants were either monetarily reimbursed for their time or received course credit. One participant was excluded from the analyses due to a self-reported diagnosis of an Axis I disorder (American Psychiatric Association, 2000). Event-related potential (ERP) analyses and a detailed description of the demographic data have been reported previously elsewhere (Oestreich et al., 2015). All participants gave written informed consent in accordance with the Declaration of Helsinki. This study was approved by the UNSW Human Research Ethics Advisory Panel (Psychology) and the University of Queensland Research Ethics Committee.

Procedure

Participants completed a number of questionnaires about their demographics, alcohol, nicotine, caffeine and recreational drug use, as well as history of Axis I disorders. Participants then underwent electroencephalographic (EEG) recordings while performing an experimental task in a quiet, dimly lit room. The experiment consisted of three conditions, namely the *Talk*, *Passive Listen* and *Cued Listen* conditions (Ford et al., 2007; Oestreich et al., 2015). Before the experiment, an instruction video was played, which demonstrated how to vocalize the syllable "ah" in a clear manner while maintaining the gaze on a fixation cross. Following the instruction video, participants were trained to vocalize the syllable "ah" with a duration of less than 300 ms and an intensity between 75 dB and 85 dB. During the *Talk* condition, participants vocalized a series of "ah"s in a desk-mounted microphone, every one to three seconds until 3 min had elapsed, producing between 75 and 125 "ahs." In the *Cued Listen* condition, participants were instructed to listen to a recording of their own willed vocalizations whilst watching a video of the vocalization waveforms. Participants were therefore able to make exact temporal predictions about the onset of a speech sound. Lastly, during the *Passive Listen* condition, participants listened to their own willed vocalizations played back without a cue. During the *Passive Listen* condition, participants were therefore unable to make temporal predictions about the onset of the next speech sound.

Of the three conditions, the *Talk* condition is distinct from the other two in that it alone involves an overt motor action. As we were interested in the functional connectivity changes associated with auditory prediction *per se*, the *Talk* condition was removed from the analysis, described below, in order to avoid the complications associated with comparing motor-active and motor-passive conditions.

Data Acquisition and Preprocessing

EEG was recorded with a 64-channel BioSemi ActiView system at a sampling rate of 2048 Hz, 18 dB/octave roll-off and

417 Hz bandwidth (3 dB). External electrodes were placed on the mastoids, the outer canthi of both eyes and below the left eye. EEG data were referenced to the average of the mastoid electrodes. Preprocessing was performed using SPM12 (Wellcome Trust Centre for Neuroimaging, London[1]) with MATLAB (MathWorks). Triggers were inserted at the onset of each "ah" and the EEG data were then segmented into 500 ms intervals with 100 ms pre- and 400 ms post-stimulus onset. Eye blinks and movements were corrected with a regression based algorithm using vertical and horizontal electrooculogram (VEOG, HEOG; Gratton et al., 1983). The low and high frequency components of the EEG signal were attenuated using a 0.5–30 Hz bandpass filter and trials containing artifacts exceeding ± 50 µV were rejected. The remaining artifact free trials were averaged per condition for each participant in order to obtain event-related potentials (ERPs). ERPs were baseline corrected using the −100 to 0 ms pre-stimulus interval. The N1 component of each ERP was defined as the most negative peak between 50 ms and 150 ms after the onset of a speech sound. In order to investigate the effect of condition on N1 amplitude at electrode Cz, a paired-samples t-test with the within-subjects factor *condition* (Passive Listen/Cued Listen) was conducted.

Dynamic Causal Modeling (DCM)

Dynamic causal modeling (DCM) relies on a generative spatiotemporal model for EEG responses evoked by experimental stimuli (Kiebel et al., 2008). It uses neural mass models (David and Friston, 2003) to infer source activity of dynamically interacting excitatory and inhibitory neuronal subpopulations (Jansen and Rit, 1995), and the connectivity established amongst different brain regions. DCM sources are interconnected via forward, backward and lateral connections (Felleman and Van Essen, 1991), and are arranged in a hierarchical manner (David et al., 2005; Kiebel et al., 2007). DCM is designed to test specific connectional hypotheses that are motivated by alternative theories (Garrido et al., 2008). Every connectivity model defines a network that attempts to predict (i.e., generate) the ERP signal. Differences in the ERPs to different experimental stimuli are modeled in terms of synaptic connectivity changes within and between cortical sources (Garrido et al., 2008). Several plausible cortical network connections are compared by estimating the probability of the data given a particular model within the space of models compared, using Bayesian Model Selection (BMS; Penny et al., 2004). BMS provides estimates of the posterior probability of the DCM parameters given the data, as well as the posterior probability of each model (Penny et al., 2004). The winning model is the model, which maximizes the fit to the data while simultaneously minimizing the complexity of the model.

The posterior probability of each model was computed over all participants using a random effects approach (RFX; Stephan et al., 2009). The conventional fixed effects approach for model comparison is limited by the assumption that all participants' data are generated by the same model and is not very robust to outliers. The RFX approach used in the current study on the other hand, is able to quantify the probability that a specific

model generated the data for any randomly chosen participant relative to other models. Moreover, RFX is robust to outliers (Stephan et al., 2009). We report the expected probability, that is, the probability that a particular model generated the data of a randomly chosen subject and the exceedance probability, which is the probability that one model is more likely than any other model, given the group data (Stephan et al., 2010). The main conclusions are based on inferences at the family level with a RFX exceedance probability of 0.95 on average (ranging from 0.85 to 1). In addition to RFX, we also report the Bayesian omnibus risk (BOR), which quantifies the risk incurred when performing Bayesian model selection, by directly measuring the probability that all model frequencies are equal (Rigoux et al., 2014). The BOR is bounded between 0 and 1, whereby a value close to 1 indicates that the models are indistinguishable, whereas a value close to 0 indicates that the models are well distinguishable from one another.

Model Specification

The models compared in this study include up to 10 brain regions hierarchically organized in one to five levels. These alternative models were motivated by speech-related brain regions that are interconnected via the auditory white matter pathways of the arcuate fasciculus and the frontal aslant. Furthermore, these brain regions have previously been reported to be activated during auditory prediction tasks similar to the paradigm used in the present study. Specifically, a study using concurrent EEG and fMRI found the superior temporal gyrus (STG), which includes Wernicke's area (W) and the primary auditory cortex (A1; Ford et al., 2016) to be activated, and a study using EEG with anatomical MRI reported activity in the STG, sensorimotor area and inferior frontal gyrus, which includes Broca's area (B; Wang et al., 2014). Since the primary auditory cortex is essential for processing auditory information, the bilateral primary auditory cortices (A1) were defined as the cortical input nodes. The arcuate fasciculus consists of a direct pathway between Wernicke's area (W) and Broca's area (B) as well as two indirect pathways, namely the posterior pathway connecting W and the Geschwind's territory (G), and the anterior pathway connecting G and B. To account for these direct and indirect connections of the arcuate fasciculus, we included models with and without G. Given the role of the frontal aslant in verbal fluency (Catani et al., 2013) and speech initiation (Fujii et al., 2016), models along the frontal aslant, which connects B with the SMA, were also included. The coordinates were chosen based on the mean Montreal Neurological Institute (MNI) coordinates for left A1 (−52, −19, 7), right A1 (50, −21, 7), left W (−57, −20, 1), right W (54, −19, 1), left G (−53, −32, 33), right G (51, −33, 34), left B (−48, 13, 17), right B (49, 12, 17), left SMA (−28, −2, 52) and right SMA (28, −1, 51; see **Figure 1**).

Since the effective connectivity associated with the prediction of external speech sounds has not been studied before, we considered a comprehensive model space including a total of 96 models comprising symmetric and non-symmetric hierarchical models, with forward (bottom-up) connections only and combined forward (bottom-up)

FIGURE 1 | Mean locations for the dynamic causal modeling (DCM) nodes and model space. The montreal neurological institute (MNI) coordinates include: left A1 (−52, −19, 7), right A1 (50, −21, 7), left W (−57, −20, 1), right W (54, −19, 1), left G (−53, −32, 33), right G (51, −33, 34), left B (−48, 13, 17), right B (49, 12, 17), left supplementary motor area (SMA; −28, −2, 52), SMA (28, −1, 51). The 48 represented models were included twice, once with *forward* connections only and once with *forward and backward* connections. These 96 models were chosen to test different hypotheses about the functional anatomy of predictability to temporally cued speech. The models were combined into five families including a *Null family*, the *Arcuate direct pathway family*, the *Arcuate direct and indirect pathways family*, the *Arcuate-Aslant direct pathways family* and the *Arcuate-Aslant direct and indirect pathways family*.

and backward (top-down) connections, with and without indirect connections between W and B via G, as well as models with and without connections along the frontal aslant, which connects B to SMA (for a full description of the model space see **Figure 1**). All models allowed for changes of intrinsic connectivity at the level of A1 and were estimated and individually compared to each other using BMS. The 96 models were then partitioned into a number of different families.

We investigated whether the prediction of external, predictable sounds is driven by feedback loops, through both forward and backward connections, or by bottom-up inputs alone, via forward connections between brain regions along the arcuate fasciculus, and possibly also through the frontal aslant. Models with feedback loops would support the predictive coding framework whereby internal predictive models are constantly updated by prediction errors resulting from the mismatch between predicted and actual auditory sensations. To this end, a family consisting of all 48 models with forward connections (i.e., *Forward family*) only was compared to a family consisting of all 48 models with forward and backward connections (i.e., *Forward and Backward family*).

Models were then grouped into families that included specific regions defined along auditory white matter tracts as follows: (1) the *Null family* consisted of eight models that included A1 only and models connecting A1 to W; (2) the *Arcuate direct pathway family* included 10 models, with connections between A1 and W as well as W and B; (3) the *Arcuate direct and indirect pathways family* consisted of 28 models including connections between A1 and W, W and G, G and B, as well as W and B; (4) the *Arcuate-Aslant direct pathways family* included 14 models with connections between A1 and W, W and B, as well as B and SMA; and (5) the *Arcuate-Aslant direct and indirect pathways family* comprising 18 models, including connections between A1 and W, W and G, G and B, W and B as well as B and SMA (see **Figures 1**, **2**).

To follow up whether models with or without the frontal aslant (i.e., connections to SMA) better explained speech sound prediction, we first combined the *Arcuate direct pathway family* (10 models with connections linking A1, W and B directly; see **Figures 1**, **2**) and the *Arcuate direct and indirect pathways family* (28 models linking A1, W, G and B) into one single family—the *Arcuate family*. We then compared this to the *Arcuate-Aslant family*, which resulted from combining the *Arcuate-Aslant direct pathways family* (14 models) and the *Arcuate-Aslant direct and indirect pathways families* (36 models) consisting of all the 50 models with connections to SMA (see **Figures 1**, **2**).

Lastly, to investigate whether Geschwind's territory is part of the circuit engaged in speech sound prediction, we compared families of models with and without connections to Geschwind's territory. To this end, we combined all models excluding Geschwind into one family—*no Geschwind family*—by grouping the *Arcuate direct pathway family* (10 models) and the *Arcuate-Aslant direct pathways family* (14 models; see **Figures 1**, **2**). We compared then the *no Geschwind family* to the *Geschwind family*, which included a combination of the *Arcuate direct and indirect*
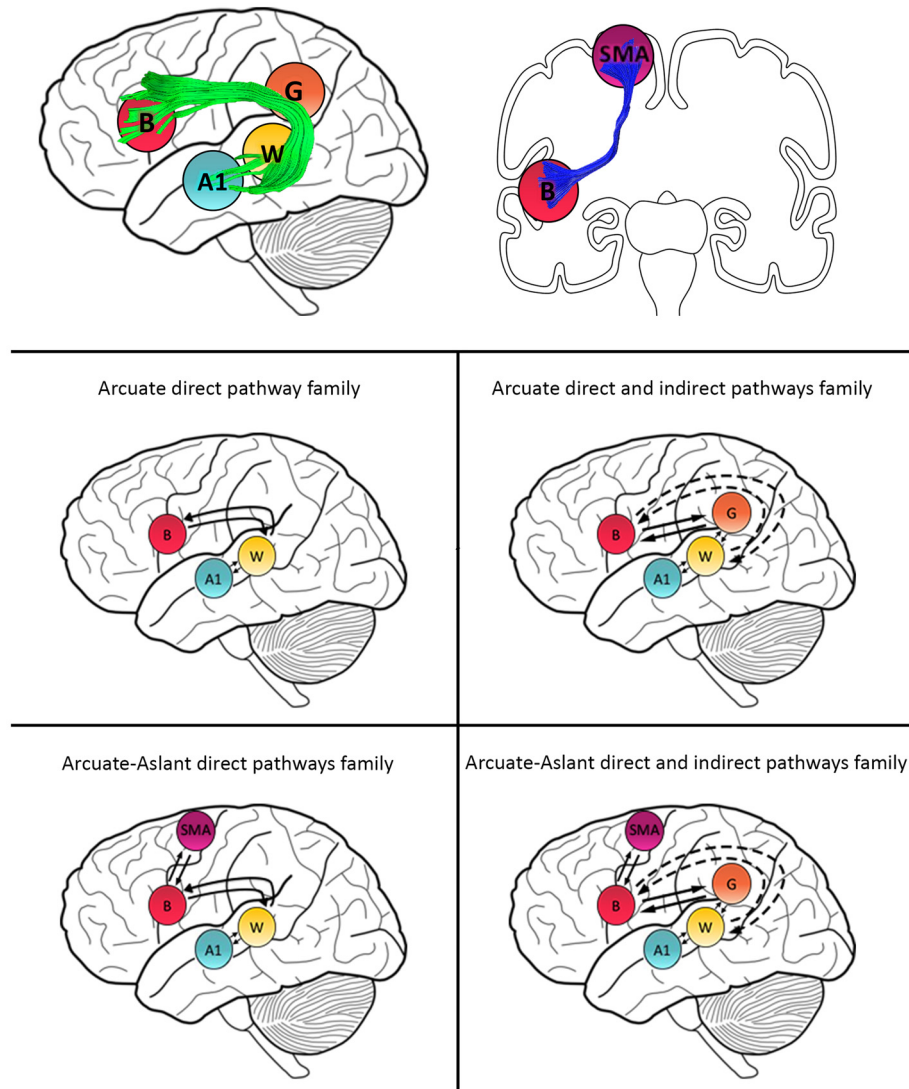
FIGURE 2 | Schematic representation of family definitions and anatomical white matter pathways. Primary auditory cortex (A1), Wernicke's area (W), Geschwind's territory (G) and Broca's area (B) are interconnected via the arcuate fasciculus (green). B and SMA are interconnected by the frontal aslant (blue).

pathways family (28 models) and the *Arcuate-Aslant direct and indirect pathways family*, that is, all the models that included Geschwind's territory (36 models). Each of the 96 models was fitted to each individual participant's mean response for the contrast between the *Passive Listen* and *Cued Listen* conditions, whereby the *Passive Listen* condition was used as the baseline condition.

RESULTS

Scalp Analysis

A paired-samples *t*-test revealed a significant difference between the *Passive Listen* and *Cued Listen* conditions on the N1-amplitude at electrode Cz ($t_{(72)}$ = 2.460, p = 0.016, *Cohen's d* = 0.288; see **Figure 3**).

DCM Analyses

In a first step all 96 models with forward (bottom-up) connections only as well as forward (bottom-up) and backward (top-down) connections were individually compared to each other. Results indicated that the best model included recurrent connections linking A1, W, G and B, as well as direct connection between W and B in both the left and right hemispheres (exceedance probability = 0.32; BOR < 0.01; see **Figure 4**). The second-best model, which was also relatively probable, was equal to the winning model except that it did include connections to SMA via the aslant in the left hemisphere (exceedance probability = 0.17; see **Figure 4**).

When comparing a family with modulations of forward (bottom-up) connections only (i.e., *Forward family*) to a family of both forward (bottom-up) and backward (top-down)

FIGURE 3 | Event-related potentials (ERPs) from electrode Cz in response to willed vocalization in the Cued Listen (magenta) and Passive Listen (cyan) conditions.
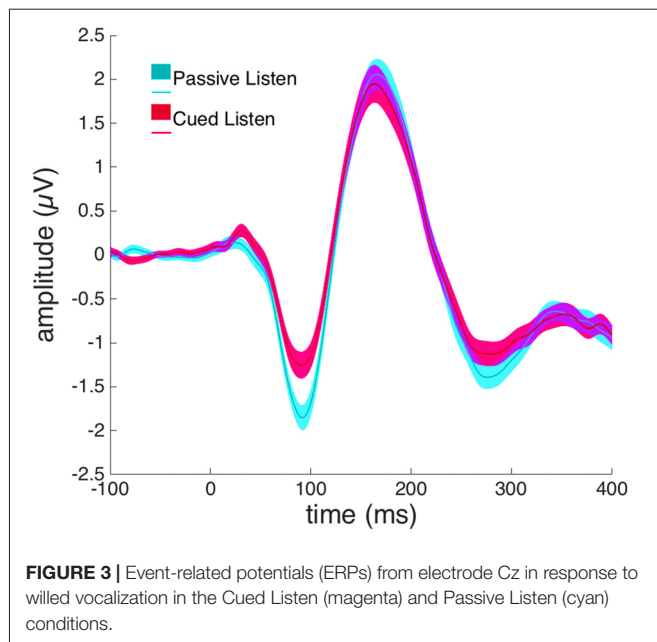
connections (i.e., *Forward and Backward family*), we found that the family consisting of a combination of *Forward and Backward* connections (expected probability = 0.56, exceedance probability = 0.85) better explained speech sound prediction than the families including *Forward* connections only.

To test specific hypotheses as to which brain regions that are interconnected by the arcuate fasciculus and the

FIGURE 4 | Model exceedance probability for attenuation of predictable speech. Bayesian model selection (random effects) over the whole model space indicated speech sound prediction was best explained by a model with recurrent (i.e., forward and backward) connections between bilateral primary auditory cortex (A1), Wernicke's area (W), Geschwind's territory (G) and Broca's area (B), as well as direct bilateral connections between W and B. This model was followed by a model, which was in all equal to the winning model except that it included a connection from B to SMA in the left hemisphere.

frontal aslant were engaged during the prediction of external, temporally-predictable speech sounds, five families of models were compared as described in the methods section (see **Figures 1**, **2**). BMS of these families indicated that the *Arcuate-Aslant direct and indirect pathways family* was the winning family (expected probability = 0.54, exceedance probability = 0.98; see **Figure 5**).

When comparing families with the arcuate fasciculus alone (i.e., *Arcuate family*) to families including both the arcuate fasciculus and the frontal aslant (i.e., *Arcuate-Aslant family*), BMS revealed that the winning, *Arcuate-Aslant family* was much more likely than the *Arcuate family* (expected probability = 0.60, exceedance probability = 0.95; see **Figure 5**).

Lastly, we investigated families of models with and without Geschwind's territory, which enquired as to whether Geschwind's territory plays a role in the functional circuit engaged in speech sound prediction (*Geschwind family* vs. *no Geschwind family*). Results indicated that the family of models including connections to Geschwind's territory outperformed the family of models without Geschwind's territory (expected probability = 0.88, exceedance probability = 1; see **Figure 5**).

DISCUSSION

This study investigated the functional anatomy underlying temporally predictable speech sounds using DCM. Model comparison revealed that modulations with both forward (bottom-up) and backward (top-down) connections better explained speech sound prediction than forward (bottom-up) connections alone. Connectivity models linking primary auditory cortex, Wernicke's area, Geschwind's territory and Broca's area via the arcuate fasciculus and the SMA, through the frontal aslant tract, outperformed models without connections to the SMA and Geschwind's territory. These findings indicate that the circuitry underlying the prediction of temporally predictable, external sounds may involve brain regions involved in the prediction of willed speech, and may include both, the arcuate fasciculus and the frontal aslant.

The finding that a combination of forward (bottom-up) and backward (top-down) connections better explained the results than forward (bottom-up) connections alone is in line with the predictive coding account, whereby a prediction is conveyed through backward (top-down) connections. Forward connections can be conceptualized as bottom-up processes (Friston, 2005; Chen et al., 2009), which convey environmental sensory information from the primary auditory cortex to higher cortical levels. On the contrary, backward connections represent top-down (Chen et al., 2009), predictive processes based on self-monitoring or past experiences. In this study, we used a *Passive Listen* condition whereby participants were passively listening to a series of previously recorded vocalizations. We used this condition as a baseline and compared it to a *Cued Listen* condition, whereby participants were cued to the exact onset of each speech sound. Therefore, participants were able to make temporal predictions about the exact onset of each speech sound, which may have

FIGURE 5 | Family-level inference for attenuation of predictable speech sounds—exceedance probabilities for the family comparisons. **(A)** Comparison of the *Forward family* (48 models) to the *Forward and Backward family* (48 models). **(B)** Comparison of five families including a *Null family* (eight models), the *Arcuate direct pathway family* (10 models), the *Arcuate direct and indirect pathways family* (28 models), the *Arcuate-Aslant direct pathways family* (14 models), and the *Arcuate-Aslant direct and indirect pathways family* (18 models). **(C)** Comparison of the *Arcuate family* (38 models) and the *Arcuate-Aslant family* (50 models). **(D)** Comparison of *no Geschwind family* (24 models) to *Geschwind family* (64 models).

been transmitted through top-down, or backward connections along the arcuate fasciculus. On the contrary, during the *Passive Listen* condition, participants were unable to make

temporal predictions about to the onset of the external sounds.

In line with these findings, Hickok (2013) proposed that the rapidity of production and comprehension of human dialog is only possible through predictive mechanisms, whereby listeners covertly imitate speakers based on their own internal representation of an utterance via top-down connections. This enables the listener to predict what the speaker is likely to say next. This theory is supported by the findings from this study whereby changes in effective connectivity from the *Passive Listen* condition to the *Cued Listen* condition are best explained by a feedback loop comprising conjoint forward (bottom-up) and backward (top-down) connections.

Another key finding of this study is that a family of models including brain areas and connections along the arcuate fasciculus (linking primary auditory cortex to Wernicke's area and Broca's area directly, and indirectly via Geschwind's territory) and the frontal aslant (connecting Broca's area directly to the SMA) best explained the prediction of temporally predictable, externally-presented speech sounds. When comparing all individual models, the winning model included connections along the arcuate fasciculus bilaterally. The second most probable model included additional connections to the frontal aslant in the left hemisphere, but only connections along the arcuate fasciculus in the right hemisphere. In order to determine whether the frontal aslant adds to the functional anatomy of speech sound prediction or whether connections along the arcuate fasciculus alone are sufficient, we compared families of all models with and without connections along the frontal aslant (while keeping the arcuate fasciculus pathways intact). The findings indicated that models with connections along the arcuate fasciculus and the frontal aslant better explained speech sound prediction than models including the arcuate fasciculus only. It may appear surprising that the family of models including the frontal aslant best explained sound prediction as the frontal aslant is thought to transmit the motor act of speech production and the *Cued Listen* condition did not involve a motor act. A possible explanation for the involvement of connections to the SMA and therefore the frontal aslant is a proposal put forward by Jackson (1958): since internal models of auditory predictions work reliably during processes of sensory motor control, the same internal models of auditory predictions, developed later in evolution, might also be utilized during higher cognitive processes such as thought or inner speech, which can be seen as the most complex motor act without actions. In the context of the present study, while participants were not actively generating the vocalization, watching the waveforms of the speech sounds might lead them to internally simulate the next vocalization, which might explain the activation of the SMA without a motor act. However, we acknowledge that this explanation is highly speculative, and should be treated with caution until supporting evidence is provided.

The arcuate fasciculus consists of long distance fibers which connect Broca's and Wernicke's area as well as short distance fibers which connect Broca's and Geschwind's territory via an anterior pathway, and Geschwind's territory and Wernicke's area via a posterior pathway (Catani et al., 2005). The results of

the present study indicate that models including long distance connections in addition to short distance connections, via Geschwind's territory, better explained sound prediction than models including long distance connections only. The direct, long distance pathway is thought to be involved in phonological repetitions (Catani and ffytche, 2005) and therefore represents a plausible connection to be utilized during this experimental tasks, whereby the same sound (i.e., a speech fragment) was played repetitively. The indirect, short distance pathways of the arcuate fasciculus are thought to be involved in semantic functions (Catani and ffytche, 2005). The engagement of these connections during the prediction of externally-presented speech sounds might be explained by the nature of the speech sounds used in the present study. Since phonemes are the building blocks of language which are used to distinguish one word from another, it is possible that participants assigned semantic meaning to these sounds, which would likely not occur if the sounds were simple tones.

The involvement of brain areas interconnected via the arcuate fasciculus during the prediction of externally-presented speech sounds is in line with findings from studies of speech sound prediction in schizophrenia. There is substantial evidence that patients with schizophrenia possess disrupted predictive coding mechanisms to self-generated speech (Ford et al., 2001, 2007; Ford and Mathalon, 2004), button-press elicited sounds (Whitford et al., 2011; Ford et al., 2014), and temporally cued sounds (Ford et al., 2007). Individuals at high-risk for developing a psychotic disorder show auditory predictive coding that is intermediate between healthy participants and patients with schizophrenia (Perez et al., 2012) and healthy individuals with psychotic-like experiences show reduced auditory predictive coding mechanisms compared to healthy individuals without psychotic-like experiences (Oestreich et al., 2015, 2016).

The mechanisms underlying these speech sound prediction deficits in schizophrenia and psychosis are still unclear. However, several studies have reported changes to the white matter structure, and specifically to the myelin sheath, of the axons constituting the arcuate fasciculus in patients with schizophrenia (Kubicki et al., 2005; Uranova et al., 2007). This is important insofar as it indicates that connectivity along the arcuate fasciculus during speech sound prediction should be delayed due to a loss of conduction velocity induced by demyelination. Support for this contention comes from a study by Whitford et al. (2011), which reported that auditory prediction abnormalities typically exhibited by patients with schizophrenia could be completely eliminated by imposing a 50 ms delay between a self-generated button press and the delivery of a sound. This was interpreted to indicate that the predictions of sensory consequences resulting from the motor command, travelling along the arcuate fasciculus during auditory prediction, were delayed by 50 ms in the group of schizophrenia patients. Furthermore, the study reported that the degree to which auditory prediction improved as a result of the delay between button press and tone delivery was linearly correlated with white matter abnormalities in the arcuate fasciculus. Furthermore, a recent study reported that predictive coding mechanisms were also disrupted in early illness schizophrenia and clinical high-risk for psychosis individuals and that the level of predictive coding abnormalities was linearly related to the microstructure of the arcuate fasciculus (Whitford et al., 2017). The findings from the present study add further support for the role of the arcuate fasciculus during auditory predictions—in this case, in the prediction of temporally predicable, but externally-generated sounds—by showing that the brain regions that are structurally interconnected by the arcuate fasciculus are also effectively connected.

DCM presents some limitations, most notably, the number of alternative models likely to explain a dataset can be very large and as a consequence, the best model might be missed if the model space is not comprehensive enough (Lohmann et al., 2012). While this is true indeed for any modeling approach that performs exhaustive searches, the objective of DCM is to perform comparisons on theoretically motivated mechanistic accounts for a given brain process. The output of DCM is the computation of an estimate for the relative evidence of different models as well as estimates about model features (i.e., connectivity parameters), rather than the specification of the single best model, which would generally have a rather small relative evidence in a large model space (Friston et al., 2013). However, to date, DCM is the only approach that integrates biophysical models of dynamic neural networks into statistical tools to investigate neuroscientific questions.

In this article we have inverted a large number of models that provided alternative mechanistic explanation for our data. Friston et al. (2016) recently introduced a new method for the analysis of group level DCM studies, which enables model selection while eschewing the need to invert all models explicitly. This approach uses parametric empirical Bayes (PEB) and Bayesian Model Reduction (BMR) to compute the posterior densities over all model parameters, under new prior densities without inverting the model again. Friston et al. (2016) demonstrated that PEB may improve the accuracy of the parameter estimates. We suggest that the use of PEB as an alternative analysis approach and a replication of this study with alternative methods to infer functional connectivity from multichannel neural EEG signals, such as phase synchronization analyses (Junfeng et al., 2012) represent fruitful avenues for future research.

In summary, the present study showed that auditory prediction to externally generated speech sounds involve brain regions such as Wernicke's area, Broca's area and Geschwind's territory, interconnected through the arcuate fasciculus via both short- and long-distance fibers, as well as the SMA, which is linked to Broca's area via the frontal aslant. Critically, we found that the prediction of externally-generated speech sounds engaged feedback loops with conjoint forward (bottom-up) and backward (top-down) connections. This result is consistent with a predictive coding framework, in which predictions are generated in higher cortical areas and communicated to lower sensory areas via backward, or top-down connections. These results also suggest that passively listening to temporally-predictable speech sounds may lead to the production of inner

speech and may engage predictions such as those believed to be involved in the production of overt speech.

AUTHOR CONTRIBUTIONS

LO designed the study, collected the data, undertook the literature review, performed the analyses and wrote the first draft of the manuscript. TW obtained funding and designed the study. MG helped with the data analysis and interpretation of results. All authors contributed to and have approved the final manuscript.

FUNDING

REFERENCES

American Psychiatric Association (2000). *Diagnostic and Statistical Manual of Mental Disorders*. Washington, DC: American Psychiatric Association Publishing.

Catani, M., and ffytche, D. H. (2005). The rises and falls of disconnection syndromes. *Brain* 128, 2224–2239. doi: 10.1093/brain/awh622

Catani, M., Dell'acqua, F., Vergani, F., Malik, F., Hodge, H., Roy, P., et al. (2012). Short frontal lobe connections of the human brain. *Cortex* 48, 273–291. doi: 10.1016/j.cortex.2011.12.001

Catani, M., Jones, D. K., and ffytche, D. H. (2005). Perisylvian language networks of the human brain. *Ann. Neurol.* 57, 8–16. doi: 10.1002/ana.20319

Catani, M., Mesulam, M. M., Jakobsen, E., Malik, F., Martersteck, A., Wieneke, C., et al. (2013). A novel frontal pathway underlies verbal fluency in primary progressive aphasia. *Brain* 136, 2619–2628. doi: 10.1093/brain/awt163

Chen, C. C., Henson, R. N., Stephan, K. E., Kilner, J. M., and Friston, K. J. (2009). Forward and backward connections in the brain: a DCM study of functional asymmetries. *Neuroimage* 45, 453–462. doi: 10.1016/j.neuroimage.2008.12.041

David, O., and Friston, K. J. (2003). A neural mass model for MEG/EEG: coupling and neuronal dynamics. *Neuroimage* 20, 1743–1755. doi: 10.1016/j.neuroimage.2003.07.015

David, O., Harrison, L., and Friston, K. J. (2005). Modelling event-related responses in the brain. *Neuroimage* 25, 756–770. doi: 10.1016/j.neuroimage.2004.12.030

Felleman, D. J., and Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex* 1, 1–47. doi: 10.1093/cercor/1.1.1

Ford, J. M., and Mathalon, D. H. (2004). Electrophysiological evidence of corollary discharge dysfunction in schizophrenia during talking and thinking. *J. Psychiatr. Res.* 38, 37–46. doi: 10.1016/s0022-3956(03)00095-5

Ford, J. M., Gray, M., Faustman, W. O., Roach, B. J., and Mathalon, D. H. (2007). Dissecting corollary discharge dysfunction in schizophrenia. *Psychophysiology* 44, 522–529. doi: 10.1111/j.1469-8986.2007.00533.x

Ford, J. M., Mathalon, D. H., Heinks, T., Kalba, S., Faustman, W. O., and Roth, W. T. (2001). Neurophysiological evidence of corollary discharge dysfunction in schizophrenia. *Am. J. Psychiatry* 158, 2069–2071. doi: 10.1176/appi.ajp.158.12.2069

Ford, J. M., Palzes, V. A., Roach, B. J., and Mathalon, D. H. (2014). Did I do that? Abnormal predictive processes in schizophrenia when button pressing to deliver a tone. *Schizophr. Bull.* 40, 804–812. doi: 10.1093/schbul/sbt072

Ford, J. M., Roach, B. J., Palzes, V. A., and Mathalon, D. H. (2016). Using concurrent EEG and fMRI to probe the state of the brain in schizophrenia. *Neuroimage. Clin.* 12, 429–441. doi: 10.1016/j.nicl.2016.08.009

Friston, K. J. (2005). A theory of cortical responses. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 360, 815–836. doi: 10.1098/rstb.2005.1622

Friston, K. J., Litvak, V., Oswal, A., Razi, A., Stephan, K. E., van Wijk, B. C. M., et al. (2016). Bayesian model reduction and empirical Bayes for group (DCM) studies. *Neuroimage* 128, 413–431. doi: 10.1016/j.neuroimage.2015.11.015

Friston, K., Daunizeau, J., and Stephan, K. E. (2013). Model selection and gobbledygook: response to Lohmann et al. *Neuroimage* 75, 275–278; discussion 279–281. doi: 10.1016/j.neuroimage.2011.11.064

Fujii, M., Maesawa, S., Ishiai, S., Iwami, K., Futamura, M., and Saito, K. (2016). Neural basis of language: An overview of an evolving model. *Neurol. Med. Chir. Tokyo* 56, 379–386. doi: 10.2176/nmc.ra.2016-0014

Gagnepain, P., Henson, R. N., and Davis, M. H. (2012). Temporal predictive codes for spoken words in auditory cortex. *Curr. Biol.* 22, 615–621. doi: 10.1016/j.cub.2012.02.015

Garrido, M. I., Friston, K. J., Kiebel, S. J., Stephan, K. E., Baldeweg, T., and Kilner, J. M. (2008). The functional anatomy of the MMN: a DCM study of the roving paradigm. *Neuroimage* 42, 936–944. doi: 10.1016/j.neuroimage.2008.05.018

Garrido, M. I., Kilner, J. M., Kiebel, S. J., and Friston, K. J. (2007). Evoked brain responses are generated by feedback loops. *Proc. Natl. Acad. Sci. U S A* 104, 20961–20966. doi: 10.1073/pnas.0706274105

Gratton, G., Coles, M. G., and Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalogr. Clin. Neurophysiol.* 55, 468–484. doi: 10.1016/0013-4694(83)90135-9

Hickok, G. (2013). Predictive coding? Yes, but from what source?. *Behav. Brain Sci.* 36, 358–358. doi: 10.1017/S0140525X12002750

Houde, J. F., and Jordan, M. I. (1998). Sensorimotor adaptation in speech production. *Science* 279, 1213–1216. doi: 10.1126/science.279.5354.1213

Jackson, J. H. (1958). *Selected Writings of John Hughlings Jackson*. New York, NY: Basic Books.

Jansen, B. H., and Rit, V. G. (1995). Electroencephalogram and visual evoked potential generation in a mathematical model of coupled cortical columns. *Biological Cybernetics.* 73, 357–366.doi: 10.1007/s004220050191

Junfeng, S., Xiangfei, H., and Shanbao, T. (2012). Phase synchronization analysis of EEG signals: an evaluation based on surrogate tests. *IEEE Trans. Biomed. Eng.* 59, 2254–2263. doi: 10.1109/tbme.2012.2199490

Kiebel, S. J., Garrido, M. I., and Friston, K. J. (2007). Dynamic causal modelling of evoked responses: the role of intrinsic connections. *Neuroimage* 36, 332–345. doi: 10.1016/j.neuroimage.2007.02.046

Kiebel, S. J., Garrido, M. I., Moran, R. J., and Friston, K. J. (2008). Dynamic causal modelling for EEG and MEG. *Cogn. Neurodyn.* 2, 121–136. doi: 10.1007/s11571-008-9038-0

Kubicki, M., Park, H., Westin, C. F., Nestor, P. G., Mulkern, R. V., Maier, S. E., et al. (2005). DTI and MTR abnormalities in schizophrenia: analysis of white matter integrity. *Neuroimage* 26, 1109–1118. doi: 10.1016/j.neuroimage.2005.03.026

Lohmann, G., Erfurth, K., Muller, K., and Turner, R. (2012). Critical comments on dynamic causal modelling. *Neuroimage* 59, 2322–2329. doi: 10.1016/j.neuroimage.2011.09.025

Mumford, D. (1991). On the computational architecture of the neocortex. I. The role of the thalamo-cortical loop. *Biol. Cybern.* 65, 135–145.

Mumford, D. (1992). On the computational architecture of the neocortex. II. The role of cortico-cortical loops. *Biol. Cybern.* 66, 241–251.

Oestreich, L. K. L., Mifsud, N. G., Ford, J. M., Roach, B. J., Mathalon, D. H., and Whitford, T. J. (2015). Subnormal sensory attenuation to self-generated speech in schizotypy: Electrophysiological evidence for a 'continuum of psychosis'. *Int. J. Psychophysiol.* 97, 131–138. doi: 10.1016/j.ijpsycho.2015.05.014

Oestreich, L. K. L., Mifsud, N. G., Ford, J. M., Roach, B. J., Mathalon, D. H., and Whitford, T. J. (2016). Cortical suppression to delayed self-initiated auditory stimuli in schizotypy: neurophysiological evidence for a continuum of psychosis. *Clin. EEG Neurosci.* 47, 3–10. doi: 10.1177/1550059415581708

Penny, W. D., Stephan, K. E., Mechelli, A., and Friston, K. J. (2004). Comparing dynamic causal models. *Neuroimage* 22, 1157–1172. doi: 10.1016/s1053-8119(04)00164-8

Perez, V. B., Ford, J. M., Roach, B. J., Loewy, R. L., Stuart, B. K., Vinogradov, S., et al. (2012). Auditory cortex responsiveness during talking and listening: early illness schizophrenia and patients at clinical high-risk for psychosis. *Schizophrenia Bulletin.* 38, 1216–1224. doi: 10.1093/schbul/sbr124

Rao, R. P., and Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* 2, 79–87. doi: 10.1038/4580

Rigoux, L., Stephan, K. E., Friston, K. J., and Daunizeau, J. (2014). Bayesian model selection for group studies—revisite. *Neuroimage* 84, 971–985. doi: 10.1016/j.neuroimage.2013.08.065

Stephan, K. E., Penny, W. D., Daunizeau, J., Moran, R. J., and Friston, K. J. (2009). Bayesian model selection for group studies. *Neuroimage* 46, 1004–1017. doi: 10.1016/j.neuroimage.2009.03.025

Stephan, K. E., Penny, W. D., Moran, R. J., den Ouden, H. E. M., Daunizeau, J., and Friston, K. J. (2010). Ten simple rules for dynamic causal modeling. *Neuroimage* 49, 3099–3109. doi: 10.1016/j.neuroimage.2009.11.015

Uranova, N. A., Vostrikov, V. M., Vikhreva, O. V., Zimina, I. S., Kolomeets, N. S., and Orlovskaya, D. D. (2007). The role of oligodendrocyte pathology in schizophrenia. *Int. J. Neuropsychopharmacol.* 10, 537–545. doi: 10.1017/S1461145707007626

Wang, J., Mathalon, D. H., Roach, B. J., Reilly, J., Keedy, S. K., Sweeney, J. A., et al. (2014). Action planning and predictive coding when speaking. *Neuroimage* 91, 91–98. doi: 10.1111/luts.12078

Whitford, T. J., Mathalon, D. H., Shenton, M. E., Roach, B. J., Bammer, R., Adcock, R. A., et al. (2011). Electrophysiological and diffusion tensor imaging evidence of delayed corollary discharges in patients with schizophrenia. *Psychol. Med.* 41, 959–969. doi: 10.1017/s0033291710001376

Whitford, T. J., Oestreich, L. K. L., Ford, J. M., Roach, B. J., Loewy, R. L., Stuart, B. K., et al. (2017). Deficits in cortical suppression during vocalization are associated With structural abnormalities in the arcuate fasciculus in early illness schizophrenia and clinical high risk for psychosis. *Schizophr. Bull.* doi: 10.1093/schbul/sbx144 [Epub ahead of print].

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Functional Neuronal Topography: A Statistical Approach to Micro Mapping Neuronal Location

Angela Jacques[1,2,3], Alison Wright[4], Nicholas Chaaya[1,2,3], Anne Overell[1,2,3], Hadley C. Bergstrom[5], Craig McDonald[6], Andrew R. Battle[1,2,7,8] and Luke R. Johnson[1,2,3,9]*

[1] Translational Research Institute, Brisbane, QLD, Australia, [2] Institute for Biomedical Innovation, Queensland University of Technology, Brisbane, QLD, Australia, [3] School of Psychology and Counseling, Queensland University of Technology, Brisbane, QLD, Australia, [4] Faculty of Health Science and Medicine, Bond University, Gold Coast, QLD, Australia, [5] Psychological Science Department, Vassar College, Poughkeepsie, NY, United States, [6] Department of Psychology, George Mason University, Fairfax, VA, United States, [7] School of Biomedical Sciences, Queensland University of Technology, Brisbane, QLD, Australia, [8] Translational Research Institute, School of Medicine, The University of Queensland Diamantina Institute, Brisbane, QLD, Australia, [9] Department of Psychiatry and Center for the Study of Traumatic Stress, Uniformed Services University School of Medicine, Bethesda, MD, United States

In order to understand the relationship between neuronal organization and behavior, precise methods that identify and quantify functional cellular ensembles are required. This is especially true in the quest to understand the mechanisms of memory. Brain structures involved in memory formation and storage, as well as the molecular determinates of memory are well-known, however, the microanatomy of functional neuronal networks remain largely unidentified. We developed a novel approach to statistically map molecular markers in neuronal networks through quantitative topographic measurement. Brain nuclei and their subdivisions are well-defined – our approach allows for the identification of new functional micro-regions within established subdivisions. A set of analytic methods relevant for measurement of discrete neuronal data across a diverse range of brain subdivisions are presented. We provide a methodology for the measurement and quantitative comparison of functional micro-neural network activity based on immunohistochemical markers matched across individual brains using micro-binning and heat mapping within brain sub-nuclei. These techniques were applied to the measurement of different memory traces, allowing for greater understanding of the functional encoding within sub-nuclei and its behavior mediated change. These approaches can be used to understand other functional and behavioral questions, including sub-circuit organization, normal memory function and the complexities of pathology. Precise micro-mapping of functional neuronal topography provides essential data to decode network activity underlying behavior.

Keywords: microanatomy, memory, network, allocation, cluster, topography, heat maps, amygdala

INTRODUCTION

Following Cajal's identification of the neuron as the fundamental functional unit of the nervous system (López-Muñoz et al., 2006), the field of neuroscience has endeavored to understand how neurons operates in local groups (ensembles) and distributed networks to bring about behavior. Cajal (1894) proposed a theory that memory storage requires the formation of new connections between neurons in the brain. How neurons and their 1000s of synaptic connections act together to encode a memory was first conceptualized by Hebb (1949) as neuronal ensembles that both spatially and temporally act together to encode a component of the memory. Since these foundational anatomical and theoretical works, newer studies involving fluorescent imaging and electron microscopy have since provided growing evidence for the modification of neuronal synapses as a result of information storage, now known as synaptic plasticity (Kandel, 2001; Korb and Finkbeiner, 2011). Thus, at the sub-cellular level knowledge of mechanisms of memory encoding is more established, in contrast at the neuronal ensemble level memory encoding mechanisms are not yet understood. Some functional evidence for Hebbian reverberatory networks connecting ensembles of neurons (Hebb, 1949) has been identified in memory circuits (Johnson et al., 2008, 2009; Josselyn et al., 2017). However, key challenges in neuroscience remain around how neurons collectively undergo plasticity in ensembles to encode memories and behaviors. Aspects of neural ensemble activity has been demonstrated in Hippocampus (Nakamura et al., 2010) and Caudate (Barbera et al., 2016) and in Amygdala (Johnson et al., 2008, 2009; Rogerson et al., 2014; Davis and Reijmers, 2017; Josselyn et al., 2017; Josselyn and Frankland, 2018). A key challenge in the neuroscience of memory is in identifying which neurons have been allocated to the memory trace and which have not, while some progress has been made (Bergstrom et al., 2008, 2011, 2013a,b; Bergstrom and Johnson, 2014; Mayford, 2014; Rogerson et al., 2014; Frankland and Josselyn, 2015; Bergstrom, 2016; Josselyn and Frankland, 2018), new techniques and approaches for understanding microanatomy are needed. This aim can be aided by the development of methods and approaches to help reliably identify and quantify systematic topographies of neurons allocated to specific memory traces.

Here, we developed a method for topographical analysis and measurement of neurons allocated to memory traces. We have applied this method to study aspects of the neurobiological encoding of fear memory. We termed this method "neuronal topographic density mapping" and have devised it to identify and map the degree of stability within a micro-topography of neurons encoding Pavlovian fear memory across different animals undergoing fear memory acquisition or extinction. The methods, described in detail below, were developed over multiple studies, investigating the location and distribution of neurons activated in fear memory in amygdala (LeDoux et al., 2006; Haranhalli et al., 2007; Bergstrom et al., 2011, 2013a,b; Johnson et al., 2012). For illustrative purposes and to expand on the scope of these techniques, we employed a small data set drawn from the study of activity-regulated cytoskeleton-associated protein (Arc) expression in prefrontal cortex.

In our studies to date, we have investigated the micro-topography of memory using Pavlovian fear conditioning. In Pavlovian or classical fear conditioning a mild foot shock [unconditioned stimulus (US)] is temporally paired with an auditory tone or comparable visual stimuli [conditioned stimulus (CS)] (Johnson et al., 2012; Bergstrom et al., 2013a; Bergstrom and Johnson, 2014). The animal learns to associate the US with the CS and exhibits typical behaviors including freezing, typical of fear/threat behavior [described extensively by other authors (LeDoux, 2000; Fanselow and Gale, 2003; Johnson et al., 2012; Josselyn and Frankland, 2018)]. We measured neurons expressing plasticity associated proteins identified by immunocytochemistry. Other functional protein and RNA expression in neurons and glia can also be used with this approach. Differences were tested in the localization of neurons among the conditioned memory groups. We have provided a methodological approach to produce topographic neuron data from brain within precisely aligned anatomical regions. This approach enables investigation of the topographic patterns of neurons expressing plasticity associated proteins in the associative fear memory formation and its extinction. We propose that this method can also be used in the reproduction of neuronal density maps with regard to many forms of neuroscience data for example, drug treatments, stress and addiction or neurodegenerative disorders.

Our methodological approach to neuron topography, described here, provides useful advantages for localizing function across behavioral conditions. Other analysis methods to measure topography also provide useful topographic data. For example, Nakamura et al. (2010) identified that memory activated neurons formed small anatomical clusters in hippocampus during place preference formation, which was identified using a cluster analysis approach. Recent studies by Barbera et al. (2016) used measures of neuronal clustering of medium spiny neurons to predict locomotive states of behavior in mice. They reported that behavioral decoding accuracy improved using spatially distinct neural clusters over single neurons (Barbera et al., 2016).

Recent advances using *in vivo* optical methods including calcium imaging have provided a rich source of complex micro anatomical and dynamic neuronal data, including in awake behaving subjects (Ohki and Reid, 2014; Romano et al., 2017; Castanares et al., 2019). Recent analysis approaches for these data include the method developed by Romano and associates, to analyze neuronal population dynamics (Romano et al., 2017). Additional recent whole brain imaging and analysis techniques by Kim et al. (2015, 2017), who developed a spatial IEG-based mapping technique as a method to view whole-brain activity. Furthermore, whole brain mapping methods have also been developed by Vousden et al. (2015) and Renier et al. (2016). Each of these methods provide the advantage of visualizing patterns of neural activity across distributed brain networks. The creation of neuronal quantitative topographic density maps, as described here, can be used for a variety of studies to pinpoint functional microcircuits in the brain.

Using our approach to mapping and measuring topography we have characterized the microanatomy and topography of neurons involved in different phases of memory, consolidation,

reconsolidation, and extinction (LeDoux et al., 2006; Haranhalli et al., 2007; Bergstrom et al., 2008, 2011, Bergstrom et al., 2013a,b; Bergstrom and Johnson, 2014). These data have the potential to pinpoint neuronal topography patterns underlying memory encoding in the mammalian brain in normal and pathological situations (Johnson et al., 2012) and thereby facilitate current treatments for pathological memory disorders (Johnson et al., 2012). The generation of neuronal topographic density maps can be used to define and measure memory allocation within the brain.

Throughout this methodological report we provide details of the rationale, procedures and equipment needed to produce and analyze topographic neuronal data. In addition, within each methodological section we provide 'examples' from our own data in order to illustrate how the methods can be applied and used. The methodological approaches we describe here have wide applications for understanding and measuring neuronal topography. Applications include measuring the topography of neurons encoding different types of memory, different sensory stimuli, and motor behaviors.

METHODS

Data Collection: Behavioral, Tissue, and Neuron Analysis in Preparation for Topographic Investigation
Run Behavioral Models
In order to produce and analyze functional neuronal topography data linked to behavior, an appropriate behavioral model is needed. Behavioral model can include a variety of learning and memory models, addiction models, social interaction models, and other behaviors of interest. In our case we have investigated in detail Pavlovian fear conditioning.

Pavlovian fear conditioning leads to the formation of associative memories. Synaptic plasticity, dependent upon phosphorylation of extracellular signal-regulated kinase (pMAPK) has been identified as critical in the formation of these memories in the lateral amygdala (LA) and medial prefrontal cortex (mPFC) (LeDoux, 2000; Fanselow and Gale, 2003; Johnson et al., 2012; Josselyn and Frankland, 2018).

Example: The sample data set consisted of fear conditioned adult male Sprague-Dawley rats (RRID:RGD_5508397) (*n* = 40) that underwent behavioral procedures in standard Pavlovian fear conditioning chambers (Coulbourn Instruments, Allentown, PA, United States) (see **Figure 1A**). The US, a 0.6 mA foot shock with duration of 500 ms, was paired with the CS, a tone of 5 kHz and 75 dB (Digitech Professional Sound Level Meter[1], 20 s in duration, to produce an associative memory. Three pairings were presented with an average 180 s inter-trial interval with total time in box of 10 min. Standard conditioning and behavioral testing procedures were followed (LeDoux et al., 2006; Haranhalli et al., 2007; Bergstrom et al.,

2008, 2011, 2013a,b; Bergstrom and Johnson, 2014). The experimenter was blind to the experimental conditions when scoring freezing behavior, which was defined as a lack of movement except that required for respiration (LeDoux et al., 1988). Next, brains were prepared for histological analysis and measurement.

Perform Immunohistochemistry
Rats were transcardially perfused and brains were post-fixed in 4% PFA overnight then stored in 0.1 M phosphate buffered saline. Free-floating serial coronal sections (40 μm) of the mPFC and amygdala were prepared using a vibratome (M11000; Pelco easiSlicer, Ted Pella, Inc., Redding, CA, United States). Sections from the LA and prefrontal cortex were labeled for pMAPK and Arc activation using the avidin–biotin peroxidase method. Detailed immunocytochemical methods can be obtained from our previous reports (see Bergstrom et al., 2011, 2013a). Slides were scanned with an Olympus VS120 slide scanner and cropped at 2x magnification (see **Figure 1B**).

Choose Anatomical Anchor/Marker
Establishing anatomical alignment between regions of interest (ROI) is necessary for visual comparison of neuron density in neural images, for sectioning the ROI into micro regions for analysis, and for both quantitative and visual analysis of the data. Therefore, choosing an appropriate anatomical anchor is a key step. The anchor point should: (1) be a readily visible anatomical feature that is close in proximity to the ROI, (2) be stable across subjects and conditions, and (3) change shape rapidly and distinctly as the viewing plane changes, so that different planes of view can be discriminated clearly. These characteristics are identifiable microscopically and importantly can also be quantified (see **Figure 1C**).

Example: The amygdala and mPFC have been implicated in Pavlovian fear conditioning (Fanselow and Gale, 2003; Johnson et al., 2012; Lee et al., 2015). In a series of studies, we have focused on the amygdala and have used the opening of the Lateral Vertical (LV) as an anatomical anchor (LeDoux et al., 2006; Haranhalli et al., 2007; Bergstrom et al., 2008, 2011, Bergstrom et al., 2013a,b; Bergstrom and Johnson, 2014). The LV has proved a useful structure for the purpose because it meets the criteria outlined above: (1) the LV is close in proximity to the amygdala, (2) the LV changes rapidly in size along the longitudinal plane, (3) the LV is a stable anatomical feature, and (4) LV changes can be seen clearly, and measured, through the sequence of planes on which the brains were sectioned, enabling quantitative analysis of the changes section by section. In order to further demonstrate and measure the properties of the LV for landmark suitability, in addition to histological measurements, we made measurements of the LV with MRI. Here, the morphological properties of the LV, including its increase in diameter along the rostral-caudal axis, were confirmed *in vivo*, using three-dimensional T2-weighted MRI to quantify its area (Bergstrom et al., 2013a). This rapid change from rostral to caudal allows for precise quantitative section alignment from

---

[1]https://www.jaycar.com.au/pro-sound-level-meter-with-calibrator/p/QM1592
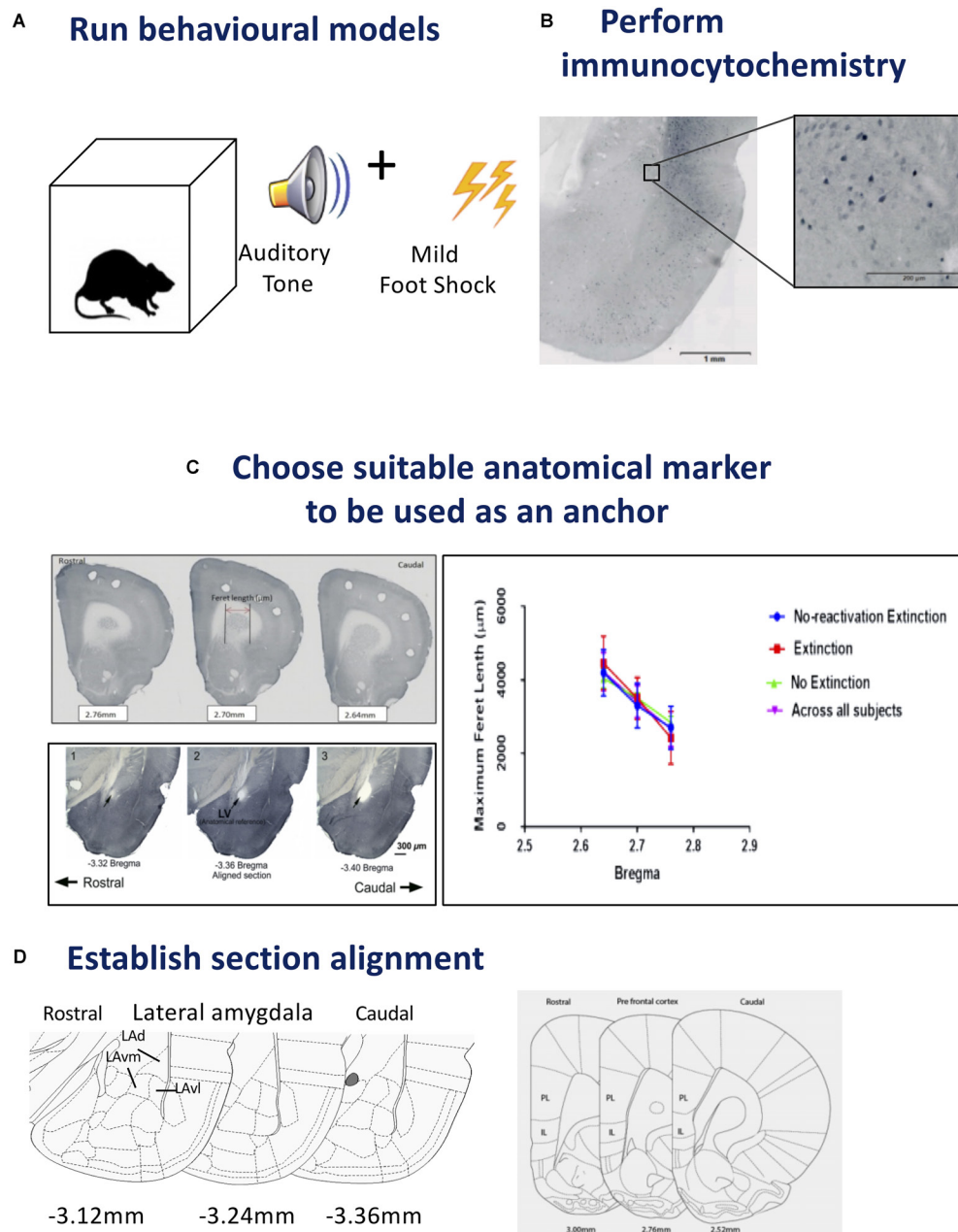
FIGURE 1 | Steps for tissue sampling and measurement from behavioral data. **(A)** *Run behavioral models*. Any expression of a chosen behavior can be used as a model. In our example we have used Auditory Pavlovian fear conditioning. Behavioral testing was conducted with adult male Sprague-Dawley rats in acoustic classical fear conditioning chambers. A 0.6 mA foot shock with duration of 500 ms was paired with a tone of 5 kHz and 75 dB, 20 s in duration to produce an associative fear memory. **(B)** *Perform immunocytochemistry*. Avidin–biotin peroxidase complex method is demonstrated here. Sections from the lateral amygdala (LA) were labeled for Arc, scanned using a slide scanner and cropped at 2x magnification. Enlarged inset square shows Arc$^+$ neurons in the dorsolateral portion of the LA at 20x magnification. Inverted gray scale images of fluorescent immunocytochemistry would also be suitable. **(C)** *Choose suitable anatomical marker to be used as an anchor*. The caudate putamen and lateral ventricle are two examples of anatomical landmarks, that we have used previously, and can be differentiated in serial sections for section alignment by Feret length within the ventile or between anatomical landmarks. Photomicrographs show three consecutive 60 μm sections across the rostrocaudal axis of the rat brain, depicting 2.76, 2.70, and 2.64 mm anterior to Bregma in the medial prefrontal cortex (mPFC). Feret diameter is shown – red arrow. Brain sections at Bregma coordinates –3.32, –3.36, and –3.40 mm posterior from Bregma were used to align the LA (Source: see Bergstrom et al., 2011). The maximum Feret length of the caudate putamen in the prefrontal cortex was shown to be statistically different across Bregma coordinates, animals and conditions. **(D)** *Establish section alignment*. The Rat Brain Atlas (Paxinos and Watson, 2007) is an important tool to assist alignment of sections. Schematic diagrams are shown depicting the regions of interest. The dorsolateral portion of the lateral amygdala (LAd), the ventromedial portion of the lateral amygdala (LAvm) and the ventrolateral portion of the amygdala (LAvl) are shown in three serial sections caudal from bregma –3.36 mm. The prelimbic (PL) and infralimbic (IL) cortex are represented by three serial sections caudal from bregma 2.52 mm. Brain Atlas diagrams are adapted from Paxinos and Watson (2007).

plane to plane. In our histological studies the morphology of the LV was reconstructed from five consecutive planes (Bregma −3.36 to −3.48). The coronal plane with the least variance between conditions was found at Bregma −3.36 in the rat (Paxinos and Watson, 2007), the entrance of the LV, so this was chosen as the most suitable anatomical anchor, in addition, it could be readily visualized and measured. At −3.36 mm Bregma, in addition to the LV it is also possible to identify the major anatomical structures of the ROI (the subnuclei of the LA). The choice of the LV as an anatomical anchor was therefore suitable because it is amygdala-centric, changes shape rapidly and clearly, and is stable across subjects (LeDoux et al., 2006; Haranhalli et al., 2007; Bergstrom et al., 2008, 2011, Bergstrom et al., 2013a,b; Bergstrom and Johnson, 2014).

We used the caudate putamen as an anatomical landmark to align sections in the prefrontal cortex (described below). Aspects of the caudate putamen met the criteria we previously set for landmark identification (see **Figure 1D**). Histological images were captured as virtual slide images (OlyVia; format.vsi) using a slide scanner (Olympus VS120). Capturing images with a slide scanner (used in this example) is an alternative approach to live capturing of neuron data with a microscope connected directly to Neurolucida as used in our previous published data (LeDoux et al., 2006; Haranhalli et al., 2007; Bergstrom et al., 2008, 2011, Bergstrom et al., 2013a,b; Bergstrom and Johnson, 2014). In this example, we used OlyVIA XV Image Viewer (Olympus Australia Pty Ltd., Vic, RRID:SCR_014342) to ascertain and measure images within a Bregma range that showed an alteration in the size of the caudate putamen. The caudate putamen becomes visible 2.7 mm anterior to Bregma, distinctly widens and lengthens in serial coronal sections across the rostrocaudal axis. Three consecutive sections (Bregma 2.7–2.58 mm) were aligned and verified across subjects and conditions by statistical comparison (ANOVA) of the Feret length (Walton, 1948) (the maximum Feret length or distance between two perpendicular tangents) was measured with Neurolucida 360 software (Neurolucida, MBF BioScience, Williston, VT, United States, RRID:SCR_001775) and analyzed with SPSS (IBM SPSS Statistics 23, WA, SCR_002865). A similar comparison of sections was calculated using z-scores from each maximum Feret measurement of the caudate putamen. No outliers were detected using ±3.0 standard deviation (SD). This principle includes 99.9% of values coming from the same normal distribution. Additionally, outliers can also be checked using online software tools, e.g., GraphPad Prism. Next, in order to test each Bregma point assignment was dissimilar and no difference existed between experimental conditions, paired t-tests were performed on the Feret measures. Each distance was found to be statistically different (example 2.76 mm Bregma; $p = 0.000304$). This data was used to help exclude misaligned sections due to natural or histological induced variations. This quantitative analysis approach can thus be used to assign sections to distinct groups maximizing alignment accuracy for subsequent neuronal topography measures.

Section Alignment

Quantitative topographical data was produced beginning with neuron identification and section alignment. While LV and caudate putamen changes can be observed through a sequence of many planes, the ROS may be rostral or caudal to this point. For this reason, the chosen landmark is used only as a point of reference. Sections are aligned manually using the landmark and working rostrally or caudally through the sequential Bregma coordinates using the measurement of width of each section as a guide. For example, Bregma 2.76 mm is 0.48 mm away from Bregma 3.24 mm; therefore, there will be 8 μm × 60 μm sections or 12 μm × 40 μm sections between the two Bregma coordinates. This highlights the need for precision when slicing and marking serial sections. Having mounted sections in the correct order on slides prior to labeling decreases time taken during this stage.

Generate Topography in Preparation for Analysis

Create Contour

In order to ensure consistency and precision in neuron counting across all subjects, a contour or tracing of the anatomical structure being investigated can be prepared in Neurolucida (NL) 360 (Neurolucida, MBF Bioscience, Williston, VT, United States). Prior to importing an image into NL for tracing, it is necessary to calibrate the image to approximate the dimensions of a single brain section bitmap image (cellSens software, Olympus, Notting Hill, VIC, Australia, RRID:SCR_014551). Within Neurolucida select > File, > Image open to allow the image to appear and select x and y calibration pixel size. These measurements are located in the image properties section in the cellSens program. Choose > Trace, > Contour Mapping in NL to begin the trace (see **Figure 2A**). The image lines may be enlarged using the zoom tool, to increase accuracy of the trace. Use the curser to trace around the selected area and > Close Contour when finished each area. This allows delineation of each section of the contour with a separate color using 'User Line.'

Scale Contour

At this point it is essential to align the contour. The size of the tracing can be adjusted to fit the image using > Tools, > Adjust Scaling. Contour alignment must be consistent across all groups, prior to neuron counting. It is advisable to open several images to scale the contour, due to minor variation in dimensions across subjects.

Calibrate Contour

Very importantly, the contour is then calibrated to a constant point (0, 0 on the x, y axis) to preserve consistency of neuron marker coordinates. The reference point is displayed by selecting > Options, > Display Preferences, > View. In this window, the radius of the point can be set to a desired diameter. Apply the display grid setting and enlarge with the magnification tools as required. The contour is moved (using move tools) such that the 0,0 coordinates are placed in the superior left corner of the contour. Once in position the contour must not move or be resized for the duration of neuron counting across all groups to

FIGURE 2 | Steps for producing raw coordinate data from identified neurons. **(A)** Create, scale, and calibrate contour. Neurolucida 360 (or an equivalent program) can be used to produce a nucleus or brain region contour from a rat brain atlas diagram. Using the contour mapping tool in Neurolucida 360 contours (in different colors) can be traced over a figure from an atlas. Lateral amygdala tracing shown was generated from Bregma −3.36 of Rat Atlas (Paxinos and Watson, 2007). **(B)** Align section to contour and mark immuno positive neurons. Prefrontal cortex section with contour overlaid. Immuno-positive neurons were marked within the contour. Saved data files can be opened in Neurolucida Explorer to gain data file information such as contour areas, Feret length measures, and neuron counts. Prelimbic contour and neurons were marked in aqua, infralimbic contour, and neurons marked in yellow. Once neurons are marked, Neurolucida Explorer (or equivalent) can import the data file to generate a contour and marker analysis, LA example shown. **(C)** Export marker coordinates. The x, y coordinates produced for each marked neuron are exported to an ASCII file which can be opened in graphing software such as Origin Pro (or equivalent). **(D)** Produce bin matrix. A data matrix is generated based on the area and density of marked neurons within the contour. Bin size is calculated using twice the area of the contour divided by the total number of neurons (De Smith et al., 2009). Once the x, y coordinates are highlighted in an Origin Pro (or equivalent) workbook, the 2D binning option under descriptive statistics is chosen. The bin ends and size can be manually entered into the dialog box once determined using the standard geospatial formula (De Smith et al., 2009).

ensure the integrity of the quantitative data. Save contour as a data file.

Align Sections to Contour

Once the tracing has been saved > CTRL + S, a scanned and cropped image of a single neural section may be opened (> File, > Image Open, > calibrate pixel size) and the tracing can be overlaid using the move tools to move only the image. There may be some minor variation in the size and properties of each subject, driven by natural variation or variations introduced during tissue processing – therefore the contour must be aligned to each section. To align the section and the contour, select > Image, > Image Processing, and > Orientation (see **Figure 2B**). Options are provided for

a mirror image, flip, 90 or 180° rotation of the image. Choose Arbitrary Rotation and use the arrows to alter the Rotation in Degrees.

Mark Immuno Positive Neurons

Once the section is aligned to the contour (or tracing), begin to mark neurons by choosing a marker from the marker toolbar located down the length of the left side of the screen. Right click the mouse button on the selected marker to rename, recolor or resize the marker. Elect to use a different color for markers in separate areas of the contour for ease of analysis at later stages of the process (see **Figure 2B**). Markers may be erased at any time during counting by > CTRL Z, or > Edit, > Undo, to remove the last placed marker. If mapping to determine the organization

of synaptic markers within the neuropile, the same procedure should be followed for marking puncta (Radley et al., 2006).

Note: If mapping neurons using NeuroLucida directly connected to a microscope for live imaging, then, following contour tracing and neuron mapping, a final alignment of all data to be compared must be made before analysis of neuron spatial distribution. Contours with mapped neurons are rotated for matched alignment using the Neurolucida Contour Alignment function.

Example: A digital image of the ROI, the mPFC, was sourced from the rat brain atlas, 6th Edn, 2007 (Paxinos and Watson, 2007, RRID:SCR_006369). Three locations, 3.3, 3.24, and 3.18 mm anterior to Bregma (Paxinos and Watson, 2007) were used for cell counting. This level was chosen as both the prelimbic and infralimbic cortices were represented at this point. Specific markers were recolored and renamed for each subregion to be mapped (**Figure 2B**).

Export Neurolucida ASCII File Into OriginPro (or Alternative)

Once all the neurons in the ROI are counted with the aligned contours, the marker coordinates (x, y, z), which Neurolucida has recorded relative to the nominated reference point, can be exported as an ASCII (plain text) file (see **Figure 2C**). To accomplish this, select > File, > Export Marker Coordinates and save the file. At this point it is also prudent to save the data file you have placed your makers on, by choosing > File, > Save Data File As. The Data file can be opened in Neurolucida Explorer > File, > open data file, > contour, > analysis, > markers and region analysis. This program provides a full synopsis of the contour areas, required for later mapping, perimeters, Feret measures, and neuron counts for each designated region. Once this information has been saved the neuron markers can be cleared in NL 360 using > Edit, > Select Objects. A window will open to the right of the screen where you can select Any Object, Only Markers, Select All, then press the Delete key. Choose > File, > Image Open to import a new section and begin the entire sequence again. Once two or more images are open, select > Image, > Image Organizer, to choose which images you will Show, Hide or Delete. Files can also be closed by selecting > File, > Close All Images. To analyze the data obtained the ASCII files can be opened in Microsoft Excel where the x and y coordinates are quickly accessed and can be cut and pasted into Origin Pro (see **Figure 1C**)[2] . Alternatively, Origin Pro has the facility to open all files at once by choosing > File, > Import, > Multiple ASCII, and following the prompts to choose the files you wish to include in one density map. It is recommended to import only files from one behavioral condition at a time to reduce human error. Once coordinates are listed, select > Descriptive Statistics, > 2D Frequency Binning, which will require input of bin sizes (Alternatives to Origin Pro can also be used – see Discussion below).

---

[2]http://www.scientificcomputing.com/product-release/2014/10/origin-and-originpro-2015-data-analysis-and-graphing-software

Select Binned Data Parameters Within Origin Pro (or Alternative)

Data binning, also known as discretization, involves grouping data into bins in order to ascertain a quantitative understanding of neuronal distribution (Kerber, 1992). Developing an appropriate data matrix relies on the optimization of the dimensions of micro regions of data (bins). This part of the analysis should be well-considered and standardized in order to closely match the bin number and dimensions with the central experimental question being investigated and also to ensure the repeatability across subjects and experiments. The number of bins can be determined based on experimenter determined parameters or alternatively a formula can be applied to standardize the selection on bin numbers and to reduce any bias in bin number selection. An established formula for this type of spatial analysis is based on twice the expected frequency of items identified in a random field (2*sampling area/n, where n = mean number of items to be counted, e.g., activated neurons) (De Smith et al., 2009). This method can be used to ensure an unbiased estimate of the optimal dimension of bins for sectioning the ROI into a matrix for data analysis. The neuron counts, and contour area measurements are obtained from the Neurolucida Explorer data. Once bin number has been calculated, the minimum bin beginning and maximum bin end for the x axis and y axis are adjusted to encompass the smallest and largest coordinates contained within the ASCII files. In Origin Pro, all Auto windows must be unchecked to allow manual input of data. The bin size is measured in micrometers squared (μm^2). Once these measurements have been entered and the number of bins is calculated by the program, select > OK (see **Figure 2D**). This converts the data into an appropriate matrix, based on the area and density of the marked objects.

Produce Bin Matrix

The next step is to use the data from the calculated matrix of bins and their corresponding neuron counts for graphing and statistical analysis. The table of bins and neurons counts derived from Origin Pro (see **Figure 2D**) can now be copied into an Excel spreadsheet (or equivalent program). Repeat this process for each ASCII file obtained from one section, in one condition across all animals – this will be based on the section alignment for a specific "Bregma" coordinate – as described above. For validation purposes individual density maps can be produced at this point, for later comparison to the mean map. For an example see a range of 26 maps produced from raw values for each subject across four experimental conditions in comparison to mean maps in Figure 2 of Bergstrom et al. (2013a).

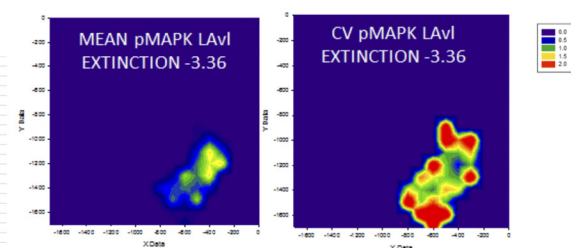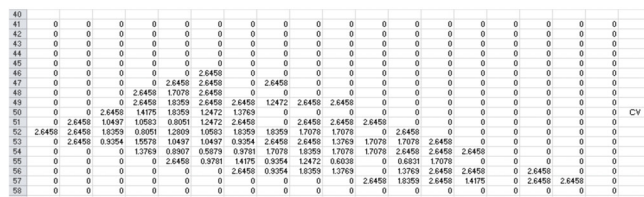## Topographic Neuronal Density Maps (Heat Maps) and Analysis
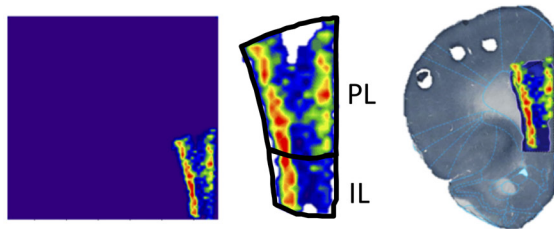### Create Density Maps

Using Excel, an average across all sheets can then be calculated – this is used to plot a graph of the mean for an experimental condition (see **Figure 3A**). In addition, from these combined and averaged data a coefficient of variance (*CV*) and other measures can be calculated. The mean and *CV* data can be used
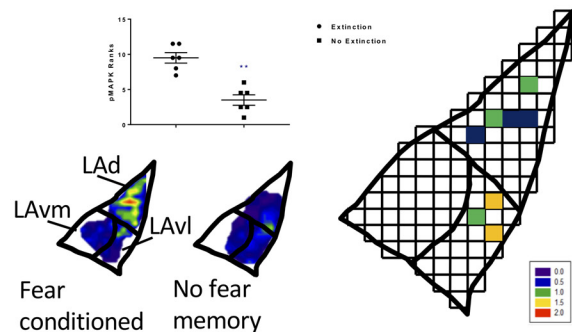
FIGURE 3 | Steps for producing and analyzing topographical density maps. **(A)** *Create topographic density map*. A neuronal topographic density mean map is produced by transferring binned data from Excel to Sigma Plot (or equivalent software) (X data = x coordinates, Y data = y coordinates). Density maps can be created for each sub region. A coefficient of variance map can be prepared by dividing the standard deviation by the mean across all samples in one condition. Difference maps can also be created between conditions. The data matrix from Origin Pro (or equivalent) is transferred to a spreadsheet. This procedure is followed for each animal from a single condition/group. An average across all sheets produces the data for a mean density map. The standard deviation is calculated and divided by the mean, producing the data required for the coefficient of variance (CV) map. Example – topographic density (mean and CV) maps shown for Bregma −3.36, pMAPK+ neurons in the ventrolateral portion of the LA of rats that underwent extinction training ($n = 7$). **(B)** *Align density map with contour and brain sections*. To enhance visualization of specific neuronal subsets, density maps can be inserted into the contours or superimposed over brain sections. Density maps may be edited to change the styles, colors, font sizes, labels etc., providing alternatives conducive to individual requirements. Information regarding cell layers can be determined from visualizing the distribution of activated neurons as shown in the pMAPK labeling of the mPFC of rats that have undergone auditory fear conditioning ($n = 7$): mean map generated in Sigma Plot (or equivalent), map placed into contour, map overlaid on rat brain section. **(C)** *Quantitative analysis of variance between conditions*. A variety of statistical analysis can be performed to compare binned data such as Bonferroni correction, principal component analysis (PCA), false discovery rate (FDR), multiple discriminant analysis and mixed model ANOVA. Example of mean maps for the expression of pMAPK in the LA provides visual comparison between auditory fear conditioned ($n = 6$) and naïve ($n = 7$) rats. pMAPK+ ranks comparing extinction ($n = 7$) and no extinction ($n = 5$) groups within the ventrolateral portion of the LA $p = 0.0022$ (*t*-test, Mann–Whitney rank and SEM).

to create separate neuron topographic density 'heat' maps using graphing software SigmaPlot or OriginPro (SigmaPlot v 12.5, Systat Software, San Jose, CA, United States RRID:SCR_003210) (or alternatives). For producing a variety of graphs from the now binned data we have used SigmaPlot, however, other programs can be used. The data matrix, using individual subject data or averaged data from Excel, is transferred beginning in the third column of SigmaPlot. The x and y coordinates from Origin Pro are copied into columns one and two of Sigma Plot. In order to produce a colored neuron topographic density 'heat' map, select > Create Graph, > Contour Tool (see **Figure 3A**). The scale can be adjusted using the graph properties tool. The production of a neuronal topographic density 'heat' map is also possible using Origin Pro.

Example: We have used bin matrix data from neurons identified and marked in the prelimbic and infralimbic cortices and transferred this data to SigmaPlot. This data was used to produce both prelimbic (PL) and infralimbic (IL) mean neuron topographic density graph (heat maps). As described above, during the creation and alignment of the contour the 0, 0 coordinate was aligned to the superior left corner of the contour. The creation of an overlay was performed by aligning this same superior left landmark of the contour with the 0,0 coordinates as displayed on the SigmaPlot contour graph export. This process allowed aligned or registered heat maps from different animals to be combined into signed maps of mean data for initial qualitative analysis of the data sets. In our example we identified neurons activated during the recall of an extinguished fear memory – initial qualitative analysis of this data reveals increased neuron density within the deep layers of the PL and IL.

Align Maps With Contours and Sections

We recommend two methods to enhance visualization of specific neuronal subsets and gain visual information regarding

distribution of activated neurons, for example, in relation to cell layer. The density maps can be inserted into the contours generated from an atlas, or alternatively superimposed over the original brain sections (see **Figure 3B**). To ensure ease of fit it is prudent to place a marker in the corner of each contour which can be removed prior to statistical analysis. Density maps may be edited in Sigma Plot to change the styles, colors, font sizes, labels etc., as requirements.

Example: Information regarding cell layers can be determined from visualizing the distribution of activated neurons as shown in the pMAPK labeling of the mPFC (see **Figure 3B**) of rats that have undergone auditory fear conditioning ($n = 7$): mean map generated in Sigma Plot (Systat Software). The map was placed into the prefrontal contour and overlaid onto a rat brain section.

Analysis of Binned Data

Graphing topographic neuron density data is an important step to provide visual evidence for changes in topography associated with behavioral and other experimental manipulations, as described above. However, when further evidence is needed to support conclusions of changes to neuronal topographic patterns then statistical analysis of the topographic data is required. Quantitative analysis can be performed with a variety of methods (discussed below) to compare topographical differences between conditions. Most common statistical software packages can be used for the analysis of topographical data. We have used GraphPad Prism 7 (GraphPad Software, Co., San Diego, CA, United States) for each of the below discussed methods, as well as linear regression and Pearson's r coefficient which can also be collected for correlation between groups.

Example: To evaluate the bins in each data matrix, two-way ANOVA with a false discovery rate (FDR) correction for multiple comparisons was conducted. The discovered bins were termed micro-regions of interest (MORIs) and assigned a color to represent the density of neuronal cell bodies located in that position (see **Figure 3C**). *Post hoc* analysis of MROIs was conducted using corrected *t*-tests.

Statistical Analysis of Topographic Neuron Density Data

In the next section, we describe statistical methods than can be applied to binned data sets of topographic data combined with behavioral manipulations to groups of experimental and control subjects. We also provide examples of application of statistical analysis from our own behavioral and neuronal topography data sets. The major challenge with the statistical analysis of multiple topographical binned data sets, combined with several experimental groups, is statistical error due to multiple comparisons. In order to best handle the analysis of topographical data we have investigated and utilized a variety of statistical approaches for large multiple comparison data sets – these include ANOVA and its variants; principal component analysis (PCA); and FDR correction (see **Table 1**). A very important step in performing statistical analysis of topographic data is to perform the statistical analysis in very close consultation with the Data produced from the topographic maps as described above. Through careful observation and consultation of the heat maps,

derived from both individual animals and importantly behavioral group mean heat maps together with their measures of variance (*CV* maps), the most meaningful analyses can be performed and interpreted.

ANOVA Followed by Bonferroni Corrected *t*-Tests

A question addressed in topographic data analysis is whether there is a significant difference in the data (e.g., number of activated neurons in the ROI) across all experimental conditions and in all ROI. One way to assess the overall difference in experimental manipulation is with analysis of variance (ANOVA), followed by a *post hoc t*-test with a correction for multiple comparisons (e.g., Bonferroni), among specific ROI and experimental groups to determine where the significance arises. Where multiple comparisons are necessary, a Bonferroni-type correction may be employed (see use in Bergstrom et al., 2011), however, it has the risk of being too strict and likely to sacrifice power in the attempt to exert stringent control over error. The potential for false negatives (type II errors) can be controlled effectively, while still retaining sufficient power, with FDR correction (Benjamini and Hochberg, 1995).

Example: We have analyzed topographic neuron density data from Pavlovian fear conditioning experiments in order to determine whether there were significant differences in topographic neuron density data across conditions by comparison of activated neuron density in each of the micro ROIs (46 bins) across all conditions via multiple comparisons (Bergstrom et al., 2013a). The mean numbers of activated neurons identified in the ROI from topographic data were used to conduct ANOVA across all conditions. Where a significant difference was found, planned contrasts between experimental and control groups were performed to assess where the differences lay (Bergstrom et al., 2013b). Multiple comparison tests involved three contrasts using one-way ANOVA. The first compared the fear conditioned and CS reactivated groups to the control groups: in this example, we compared box alone and CS (memory not reactivated groups). The second contrast was between the fear conditioned and CS reactivated groups and the third compared the box alone to the CS group. Having established a significant difference across conditions and located the main effect between experimental and control conditions, the next step was to locate the region of greatest variance in the ROI, requiring assessment of the differences in micro ROIs between groups (Bergstrom et al., 2013a). Furthermore, we also ran correlations with behavioral data as additional analysis (Bergstrom et al., 2013a).

False Discovery Rate (FDR)

Where the area under investigation has been sectioned into topographical units, each having its own data set, multiple ANOVAs on all topographical units may determine more precisely any variance between experimental conditions. FDR controls the expected rate of false rejection of the null hypothesis, by setting a parameter, the quotient q, as the "tolerable" FDR (Genovese et al., 2002). The *q*-value is used as an alternative to *p*-value when reporting significance, and while it may be set

at a conventional level (0.05), a higher level may be reasonable (Genovese et al., 2002). FDR has been used effectively in neuroscientific studies (Genovese et al., 2002; Groppe et al., 2011; Bergstrom et al., 2013a; Bergstrom and Johnson, 2014). Once the region of greatest variance across all conditions is identified, follow up tests focus the investigation on the variance between experimental conditions, in those locations.

Example: We have previously successfully applied FDR for type II error minimization and identification of significance in specific topographic ROI in behavioral experiments (see Bergstrom et al., 2013a; Bergstrom and Johnson, 2014). In these studies, we conducted mass univariate ANOVAs to assess differences in neuron activation across all conditions in each of 46 bins. FDR correction was used, with the tolerable limit set at $q = 0.1$. Significant differences across conditions were found in certain micro ROIs (nine of 46 bins), so comparisons were performed on those particular data to locate (1) the effect of the experimental versus control groups and (2) the difference between two experimental groups (Bergstrom et al., 2013a; Bergstrom and Johnson, 2014). The q-values were mapped onto the topographical matrix (bins) to reveal the highly localized topography of neuronal activation. The spatial distribution of these points of significance was confirmed on visual analysis of the neuronal topographic density maps compiled from topographic data, and also reflected earlier findings (Bergstrom et al., 2011). Subsequent correlational analysis was used to confirm the relationship between the density of marked neurons and behavior.

Principal Component Analysis (PCA)

Another approach to topographical data with multiple ROI and group comparisons is PCA. PCA seeks to identify and rank combinations of variables that account for variance within the data set. PCA enables the relationships between these patterns of variables to be identified, tested and confirmed (Jolliffe, 2002). PCA has been applied by ourselves and others to address a variety of anatomical questions, for example in morphological studies of microglial cells (Soltys et al., 2005); and vagus nerves (Horn and

Friedman, 2003); localization of sensory cells in the thalamus in facial recognition (Chapin and Nicolelis, 1999); the segregation of pyramidal neurons into morphological defined cell populations (Bergstrom et al., 2008); eye-tracking data (Bergstrom et al., 2016); and extensivley in MRI data (Lin F. et al., 2006).

Example: We have successfully applied PCA for the analysis of topographic neuronal density data activated in studies of Pavlovian fear conditioning. Activated neurons were mapped and the area sectioned into micro ROIs (bins) as described above, to produce a matrix of memory data (Bergstrom et al., 2011, 2013a). Ten components (of spatial data) were revealed, with one of these (SC1) being associated with the pattern of greatest difference (principal component score) in the spatial distribution of activated neurons between experimental conditions. SC1 displayed a unique pattern of activated neurons in a particular subnucleus of the amygdala (the LAd) across all brain samples in the experimental group. This was confirmed by t-test comparisons (Bonferroni corrected) of the bins with the most prominent loading values, and these also correlated with the area of highest density in the topographic analysis outlined above. That is, as described above, the statistical pattern could be confirmed by visual patterns seen in the neuronal topographic density maps generated by color-coding neuron densities. PCA has proved a useful statistical tool to extract meaningful patterns of variance related to the experimental manipulation, which could be confirmed by both comparison with visual representations of the data and Bonferroni corrected t-tests (Bergstrom et al., 2011, 2013a).

Multiple Discriminant Analysis (MDA)

Multiple discriminant analysis (MDA) is a method of visualizing patterns within complex data sets (Lin L. et al., 2006). With complex data, such a topographic data with many anatomical sub-regions and bins combined with multiple experimental conditions, where both location and distribution across area, are under investigation it can be important to identify patterns within this data set, in order to help understand and interpret the data. MDA can be used to determine how a set of continuous

TABLE 1 | Approaches for statistical analysis of neuron topographic data.

Method	Purpose	Advantage
ANOVA followed by Bonferroni corrected t-tests	To define where there is a significant difference in the data across conditions	Stringent control over type II errors
False discovery rate	To locate specific topographic regions of greatest variance across all conditions	Controls the expected rate of false rejection of the null hypothesis)
		Greater power
		Can be useful prior to correlational analysis
Principal component analysis	Identifies and ranks combinations of variables that account for variance within the data set	Extract meaningful patterns of neuronal variance related to the experimental manipulation
Multiple discriminant analysis	To visualize patterns within complex data sets	Determines how a set of continuous variables can discriminate groups
Mixed model ANOVA	Tests for differences between independent groups while using repeated measures to analyze topographic data combined with experimental manipulations	No adjustment for multiple comparisons is required
		Accounts for random effects
		*GEE and **GAMM can be applied after, to accommodate non-linear relationships

*GEEs, generalized estimated equations; **GAMMs, generalized additive mixed models.*

variables can discriminate groups (Bergstrom et al., 2013b), for example, how the pattern of neuron density in certain subnuclei (the independent or predictor variable) can predict the experimental condition the subject brain best fits into (the grouping or independent variable). MDA gives loading values (canonical variate correlation coefficients) that represent the relative contribution of each variable in a set of variables (a dimension) that discriminates groups from each other (see Lin L. et al., 2006; Bergstrom et al., 2013b).

Example: In one topography of Pavlovian fear memory study, we were interested in the relative contribution of lateral and basal amygdala (LA) subnuclei to the overall density of activated (pERK/MAPK expressing) neurons among each experimental condition (Bergstrom et al., 2013b). First, MANOVA was performed to examine the relationship among the subnuclei. Where a significant relationship was found, one-way ANOVA on each subnucleus tested for significant differences between conditions. Next, MDA was used to test the relative contribution of each subnucleus to the overall difference in density of activated neurons between conditions. The MDA revealed a single underlying pattern in density of activated neurons across lateral and basal amygdala subnuclei that discriminated the experimental and control groups. It also showed the subnucleus (the LAd) that contributed most to the overall difference between conditions. Having used MDA to help identify the region with the most significant contribution to the overall pattern of variance between conditions, it was possible to go further and explore more fine-grained details within the data. To confirm the pattern identified with MDA, *post hoc* comparisons with Bonferroni correction were performed, verifying the findings on the location and experimental condition of the greatest activation, and reinforcing ours and others previous findings about the predominance of LAd neural plasticity in fear memory (Rodrigues et al., 2004; Bergstrom et al., 2011).

Mixed Model ANOVA

The Mixed Model ANOVA also known as a Mixed Design ANOVA or a Split-Plot ANOVA, allows for testing for differences between independent groups (in functional topography experiment these will be the impendent behavioral groups, i.e., experiment and control groups) while using repeated measures (bins in topography experiments). Thus, the Mixed Model ANOVA can be employed for microanatomy data comprising neuron counts within bins contrasted across several independent groups. For our studies of functional neuronal topography, we typically derive 20–80 bins per animal comprising the within-group dependent variable. For the independent variable, several independent groups of animals are used including experiment and control groups. Mixed Models allow for the analysis of data from all locations and all animals in one analysis. Thus, Mixed Models a have strong potential for analysis of topographic data combined with experimental manipulations – such as behavioral or pharmacological manipulations. Using a Mixed Model analysis data between anatomic locations can be compared and no adjustment for multiple comparisons is required. Mixed Models can be thought as an advancement of ANOVA and regression

models. One, very important but often overlooked, assumption of ANOVA/Regression, is that the data are independent of each other. Thus, the analysis cannot have the same individual represented twice in the same dataset. For example, measurements on LA have to be analyzed separately from infralimbic cortex.

Mixed models ANOVA offers a toolbox to account for the dependence of measurements taken on the same individual, by accounting for, so called, *random effects.* Random effects are variables for which we are not interested in the actual levels that we have sampled but on what they represent as a sample from a population. The most usual random effect would be the individual animal (for further definitions of *random effects* readers are directed to Fitzmaurice et al. (2004) and Zuur et al. (2007, 2009). Methods related to Mixed Model ANOVA that could also be applied to topographic data sets with is the generalized estimated equations (GEEs) and the generalized additive mixed models (GAMMs) which can accommodate non-linear relationships (for further information see, Zuur and Ieno, 2016 for GAMM and Fitzmaurice et al., 2004 on Mixed Model ANOVA and GEEs and their differences).

DISCUSSION

Understanding neural network organization and predicting memory and behavior from neural network functionality is a critical goal in the field of neuroscience. Although various imaging techniques are capable of large-scale analysis of functional brain regions, they are not suitable for imaging the spatial distribution, connectivity and stability of neurons at the micro-network level. The ability to accurately map, measure and compare neural network spatial properties, as described here, contributes to our fundamental awareness of the organization and structure of functional neural circuits. Classic cellular and molecular analysis of neuronal tissue assists in the identification of molecular machinery underlying behavior but does not answer questions relative to the fundamental organizational properties and their functional changes associated with behavior. We have developed a combined topographic and statistical approach for producing and analyzing micro-topographic data. This method provides clear visualization of the spatial organization and degree of consistent neuronal patterns across brains from individual subjects and in different experimental conditions.

Neuronal material used for topographic mapping can include both exogenously labeled, such as immunocytochemistry and *in situ* hybridization, as well as endogenous genetic labeling with green fluorescent protein (GFP) and other fluorescent probes. Consistency in labeling is important with regard to whichever neuron marking system is selected for topographic mapping. The statistical methods recommended and applied here allow for natural variation in measured populations. Nonetheless, reduction of variability will improve outcome consistency and statistical verifications. Marking neurons requires consistent labeling and consistent identification

of neurons. To verify consistency, ideally experimenters blind to the experimental conditions are employed throughout or for verification checks of large data sets. The general principles outlined here for micro-topographic mapping can be applied to sectioned brain material as well as whole brain analysis approaches using CLARITY, CUBIC, or iDISCO. Three-dimensional analysis also requires focus and comparative measurements on specific anatomic ROS. Both 2D and 3D analysis ultimately requires localization and correlation of cellular activity with behavioral function using the approaches described here.

Topographic Mapping

The first step in the approach to visual and quantitative analysis of functional neuronal topography between animals is to establish section alignment. Careful choice of an appropriate and stable landmark or anchor point associated with the ROS is essential (LeDoux et al., 2006; Haranhalli et al., 2007; Bergstrom et al., 2011, 2013a,b; Johnson et al., 2012; Bergstrom and Johnson, 2014). Identification of an anchor point which has rapid and distinct conformational change through sectional view planes will ensure success at this level. The second stage involves fitting a contour to the ROS, which ensures precision of the region in which the neurons will be counted, as well as consistency in the area across subjects. A limitation at this stage is small variation between sections from each subject, which can come from animal variations and also from histological processing, therefore care is needed to minimize variation. The contour must be fitted to each section with a degree of individual judgment. Specific brain regions, such as the hippocampus, may also significantly change in shape along the longitudinal axis and therefore a single contour is not feasible. An alternate approach entails producing a unique mean contour section for a specific data set. The rat brain atlas, developed by Paxinos and Watson in the 1980s (Paxinos and Watson, 2007), is one of the most established and detailed sources of anatomical coordinates available at this time. Other brain atlases are available and can also be used. In the Paxinos and Watson atlases, the depicted brain sections can appear up to 480 micrometers apart necessitating several brain sections to be mapped to individual atlas plates. Our method is therefore limited in part by the standardized atlas information currently available (Paxinos and Watson, 2007).

Prior to creating a contour an atlas image generally requires resizing, which can represent an amount of time spent making adjustments with various software packages. Due to the number of software packages used to produce the images, it is essential to note both the accepted file types (as listed in methods above) for compatibility as images are moved between programs. Furthermore, it is very important to note the numerical functions involved in any resizing, so that consistency is maintained. Computer processing speed and memory requirements must also be considered when using the large data files produced by slide scanning.

Free, open source programs are available for some procedures, making our described method economically viable to all. For example, Image J and FIJI (National Institutes of Health) can be substituted for some elements of the topographic mapping, as it is

able to perform cell counts and export x,y coordinate data. Image J has many plugins available and runs in Java which is editable. Prior to this the contours must be calibrated to a zero point to facilitate precise individual comparisons. Once the coordinates have been exported a data matrix may be developed. Data bins are created using a geospatial analysis formula to establish unbiased bin dimensions. Open source programs are also available for this step requiring some degree of coding for specific features. QtiPlot (Free Software Foundation) is a free replacement for Origin and SigmaPlot. It will enable binning of x, y coordinates into a two-dimensional matrix and has contour generating capabilities for producing neuronal topographic density maps. Free online software for FDR analysis, as described above, is also available[3]. While we have outlined and described our methodical approach using a series of standalone commercial software packages for each of the steps descried, free software is also available making the methodical approaches described here freely available for all worldwide.

Analysis of Topographic Data

Although we have presented several arguments for the use of binned data for micro-topographic analysis, there remains the opinion that discretization has limitations (MacCallum et al., 2002; Langseth, 2008). We have used both PCA as well as Mass Univariate ANOVA with FDR correction as a useful way to locate areas of most variance in complex data, and to confirm the qualitative data from our mean heat maps. This method assists in decreasing the reduction in power generated with Bonferroni procedures (Verhoeven et al., 2005). While we provide general guidance for analysis of binned micro-anatomical data sets, we advise the reader to liaise with statisticians to evaluate the methodical approaches described here with the chosen data analysis techniques for the analysis of unique data sets and research questions.

CONCLUSION

Neuronal micro-topographic density maps can assist in defining specific brain regions involved in behavior. Statistically verified microanatomical mapping has the ability to advance our knowledge of the multi layered, complex organization of the brain and its cognitive systems. Our approach for the measurement and contrasting of neuronal topographic data in behavioral experiments has been successfully applied to the study of the microanatomy of memory formation. It has enabled us to visualize the spatial allocation of neurons activated during the acquisition of fear memories (LeDoux et al., 2006; Haranhalli et al., 2007; Bergstrom et al., 2011, 2013a,b; Johnson et al., 2012; Bergstrom and Johnson, 2014). We propose this method will prove advantageous to other forms of neuroscience, including the cellular basis of addiction; pathological memory models; pharmacological manipulations, and other forms of functional microanatomy (Johnson et al., 2012; Holmes and Singewald, 2013).

---

[3]sdmproject.com

Existing nuclei cataloged in brain atlases have been defined histologically, our approach allows for the identification of new functional micro-regions within established brain nuclei. By providing this walk-through tutorial we encourage further development of these goals.

ETHICS STATEMENT

This study was approved by The University of Queensland Animal Ethics Committee.

AUTHOR CONTRIBUTIONS

AJ wrote paper, made figures, contributed to methods development. AW contributed to methods development. NC contributed to methods development. HB contributed to methods development and development of statistical approaches. CM contributed to methods development and development of statistical approaches. AO contributed to methods development. AB co-wrote the paper. LJ conceived and developed method, co-wrote the paper.

REFERENCES

Barbera, G., Liang, B., Zhang, L., Gerfen, C. R., Culurciello, E., Chen, R., et al. (2016). Spatially compact neural clusters in the dorsal striatum encode locomotion relevant information. *Neuron* 92, 202–213. doi: 10.1016/j.neuron. 2016.08.037

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B Methodol.* 57, 289–300.

Bergstrom, H. C. (2016). The neurocircuitry of remote cued fear memory. *Neurosci. Biobehav. Rev.* 71, 409–417. doi: 10.1016/j.neubiorev.2016.09.028

Bergstrom, H. C., and Johnson, L. R. (2014). An organization of visual and auditory fear conditioning in the lateral amygdala. *Neurobiol. Learn. Mem.* 116, 1–13. doi: 10.1016/j.nlm.2014.07.008

Bergstrom, H. C., Mcdonald, C. G., Dey, S., Fernandez, G. M., and Johnson, L. R. (2013a). Neurons activated during fear memory consolidation and reconsolidation are mapped to a common and new topography in the lateral amygdala. *Brain Topogr.* 26, 468–478. doi: 10.1007/s10548-012-0266-6

Bergstrom, H. C., Mcdonald, C. G., Dey, S., Tang, H., Selwyn, R. G., and Johnson, L. R. (2013b). The structure of pavlovian fear conditioning in the amygdala. *Brain Struct. Funct.* 218, 1569–1589. doi: 10.1007/s00429-012-0478-2

Bergstrom, H. C., Mcdonald, C. G., French, H. T., and Smith, R. F. (2008). Continuous nicotine administration produces selective, age-dependent structural alteration of pyramidal neurons from prelimbic cortex. *Synapse* 62, 31–39. doi: 10.1002/syn.20467

Bergstrom, H. C., Mcdonald, C. G., and Johnson, L. R. (2011). Pavlovian fear conditioning activates a common pattern of neurons in the lateral amygdala of individual brains. *PLoS One* 6:e15698. doi: 10.1371/journal.pone.0015698

Bergstrom, J. C. R., Olmsted-Hawala, E. L., and Bergstrom, H. C. (2016). Older adults fail to see the periphery in a web site task. *Univ. Access Inf. Soc.* 15, 261–270. doi: 10.1007/s10209-014-0382-z

Cajal, S. R. Y. (1894). The croonian lecture: la fine structure des centres nerveux. *Proc. R. Soc. Lond.* 55, 444–468. doi: 10.1098/rspl.1894.0063

Castanares, M., Stuart, G. J., and Daria, V. (2019). "Holographic functional calcium imaging of neuronal circuit activity," in *Advanced Optical Methods for Brain Imaging*, eds F. J. Kao, G. Keiser, and A. Gogoi (Berlin: Springer), 143–165.

Chapin, J. K., and Nicolelis, M. A. (1999). Principal component analysis of neuronal ensemble activity reveals multidimensional somatosensory representations. *J. Neurosci. Methods* 94, 121–140. doi: 10.1016/S0165-0270(99)00130-2

Davis, P., and Reijmers, L. G. (2017). The dynamic nature of fear engrams in the basolateral amygdala. *Brain Res. Bull.* 141, 44–49. doi: 10.1016/j.brainresbull. 2017.12.004

De Smith, M. J., Goodchild, M. F., and Longley, P. A. (2009). *Geospatial Analysis: a Comprehensive Guide to Principles, Techniques and Software Tools*, 3rd Edn. Leicester: Troubador Ltd.

Fanselow, M. S., and Gale, G. D. (2003). The amygdala. fear, and memory. *Ann. N. Y. Acad. Sci.* 985, 125–134. doi: 10.1111/j.1749-6632.2003.tb07077.x

Fitzmaurice, G., Laird, N., and Ware, J. (2004). *Applied Longitudinal Analysis*. Hoboken, NY: Wiley-Interscience.

Frankland, P. W., and Josselyn, S. A. (2015). Memory allocation. *Neuropsychopharmacology* 40:243. doi: 10.1038/npp.2014.234

Genovese, C. R., Lazar, N. A., and Nichols, T. (2002). Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage* 15, 870–878. doi: 10.1006/nimg.2001.1037

Groppe, D. M., Urbach, T. P., and Kutas, M. (2011). Mass univariate analysis of event-related brain potentials/fields I: a critical tutorial review. *Psychophysiology* 48, 1711–1725. doi: 10.1111/j.1469-8986.2011.01273.x

Haranhalli, N., Massie, L., Little, J., Ledoux, J., and Johnson, L. R. (2007). "Relationship between memory 'strength' and the number of phosphorylated MAPK neurons in the lateral amygdala," in *Proceedings of the Neuroscience Meeting Planne*, San Diego, CA.

Hebb, D. O. (1949). *The Organization of Behavior: A Neuropsychological Theory*. London: Psychology Press.

Holmes, A., and Singewald, N. (2013). Individual differences in recovery from traumatic fear. *Trends Neurosci.* 36, 23–31. doi: 10.1016/j. tins.2012.11.003

Horn, C. C., and Friedman, M. I. (2003). Detection of single unit activity from the rat vagus using cluster analysis of principal components. *J. Neurosci. Methods* 122, 141–147. doi: 10.1016/S0165-0270(02)00304-7

Johnson, L. R., Hou, M., Ponce-Alvarez, A., Gribelyuk, L. M., Alphs, H. H., Albert, L., et al. (2008). A recurrent network in the lateral amygdala: a mechanism for coincidence detection. *Front. Neural Circuits* 2:3. doi: 10.3389/neuro.04.003.2008

Johnson, L. R., Ledoux, J. E., and Doyere, V. (2009). Hebbian reverberations in emotional memory micro circuits. *Front. Neurosci.* 3, 198–205. doi: 10.3389/neuro.01.027.2009

Johnson, L. R., Mcguire, J., Lazarus, R., and Palmer, A. A. (2012). Pavlovian fear memory circuits and phenotype models of PTSD. *Neuropharmacology* 62, 638–646. doi: 10.1016/j.neuropharm.2011.07.004

Jolliffe, I. (2002). *Principal Component Analysis*. New York, NY: Springer-Verlag.

Josselyn, S. A., and Frankland, P. W. (2018). Memory allocation: mechanisms and function. *Annu. Rev. Neurosci.* 41, 389–413. doi: 10.1146/annurev-neuro-080317-061956

Josselyn, S. A., Kohler, S., and Frankland, P. W. (2017). Heroes of the engram. *J. Neurosci.* 37, 4647–4657. doi: 10.1523/JNEUROSCI.0056-17.2017

Kandel, E. R. (2001). The molecular biology of memory storage: a dialogue between genes and synapses. *Science* 294, 1030–1038. doi: 10.1126/science.1067020

Kerber, R. (1992). "Chimerge: discretization of numeric attributes," in *Proceedings of the Tenth National Conference on Artificial Intelligence*. (Menlo Park, CA: AAAI Press), 123–128.

Kim, Y., Venkataraju, K. U., Pradhan, K., Mende, C., Taranda, J., Turaga, S. C., et al. (2015). Mapping social behavior-induced brain activation at cellular resolution in the mouse. *Cell Rep.* 10, 292–305. doi: 10.1016/j.celrep. 2014.12.014

Kim, Y., Yang, G. R., Pradhan, K., Venkataraju, K. U., Bota, M., Del Molino, L. C. G., et al. (2017). Brain-wide maps reveal stereotyped cell-type-based cortical architecture and subcortical sexual dimorphism. *Cell* 171, 456–469. doi: 10.1016/j.cell.2017.09.020

Korb, E., and Finkbeiner, S. (2011). Arc in synaptic plasticity: from gene to behavior. *Trends Neurosci.* 34, 591–598. doi: 10.1016/j.tins.2011. 08.007

Langseth, H. (2008). "Bayesian networks in reliability: the good, the bad, and the ugly," in *Advances in Mathematical Modeling for Reliability*, eds T. Bedford, J.

Quigley, L. Walls, B. Alkali, A. Daneshkhah, and G. Hardman (Amsterdam: IOS Press).

LeDoux, J. E. (2000). Emotion circuits in the brain. *Annu. Rev. Neurosci.* 23, 155–184. doi: 10.1146/annurev.neuro.23.1.155

LeDoux, J. E., Haranhalli, N., Massie, L. W., Little, J. P., and Johnson, L. R. (2006). Systematic constellations of neurons are activated in lateral amygdala following Pavlovian fear conditioning, Abstract number 370.8. Society for Neuroscience Abstracts.

LeDoux, J. E., Iwata, J., Cicchetti, P., and Reis, D. (1988). Different projections of the central amygdaloid nucleus mediate autonomic and behavioral correlates of conditioned fear. *J. Neurosci.* 8, 2517–2529. doi: 10.1523/JNEUROSCI.08-07-02517.1988

Lee, H. J., Haberman, R. P., Roquet, R., and Monfils, M. H. (2015). Extinction and retrieval + extinction of conditioned fear differentially activate medial prefrontal cortex and amygdala in rats. *Front. Behav. Neurosci.* 9:369. doi: 10.3389/fnbeh.2015.00369

Lin, F., Yu, C., Jiang, T., Li, K., Li, X., Qin, W., et al. (2006). Quantitative analysis along the pyramidal tract by length-normalized parameterization based on diffusion tensor tractography: application to patients with relapsing neuromyelitis optica. *Neuroimage* 33, 154–160. doi: 10.1016/j.neuroimage.2006.03.055

Lin, L., Osan, R., and Tsien, J. Z. (2006). Organizing principles of real-time memory encoding: neural clique assemblies and universal neural codes. *Trends Neurosci.* 29, 48–57. doi: 10.1016/j.tins.2005.11.004

López-ñoz, F., Boya, J., and Alamo, C. (2006). Neuron theory, the cornerstone of neuroscience, on the centenary of the Nobel Prize award to Santiago Ramón y Cajal. *Brain Res. Bull.* 70, 391–405. doi: 10.1016/j.brainresbull.2006.07.010

MacCallum, R. C., Zhang, S., Preacher, K. J., and Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychol. Methods* 7, 19–40. doi: 10.1037/1082-989X.7.1.19

Mayford, M. (2014). The search for a hippocampal engram. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 369:20130161. doi: 10.1098/rstb.2013.0161

Nakamura, N. H., Fukunaga, M., Akama, K. T., Soga, T., Ogawa, S., and Pavlides, C. (2010). Hippocampal cells encode places by forming small anatomical clusters. *Neuroscience* 166, 994–1007. doi: 10.1016/j.neuroscience.2009.12.069

Ohki, K., and Reid, R. C. (2014). In vivo two-photon calcium imaging in the visual system. *Cold Spring Harb. Protoc.* 2014, 402–416. doi: 10.1101/pdb.prot081455

Paxinos, G., and Watson, C. (2007). *The Rat Brain in Stereotaxic Coordinates. Qingchuan Zhuge Translate*. Beijing: People's Medical Publishing House, 32.

Radley, J. J., Johnson, L. R., Janssen, W. G., Martino, J., Lamprecht, R., Hof, P. R., et al. (2006). Associative Pavlovian conditioning leads to an increase in spinophilin-immunoreactive dendritic spines in the lateral amygdala. *Eur. J. Neurosci.* 24, 876–884. doi: 10.1111/j.1460-9568.2006.04962.x

Renier, N., Adams, E. L., Kirst, C., Wu, Z., Azevedo, R., Kohl, J., et al. (2016). Mapping of brain activity by automated volume analysis of immediate early genes. *Cell* 165, 1789–1802. doi: 10.1016/j.cell.2016.05.007

Rodrigues, S. M., Schafe, G. E., and Ledoux, J. E. (2004). Molecular mechanisms underlying emotional learning and memory in the lateral amygdala. *Neuron* 44, 75–91. doi: 10.1016/j.neuron.2004.09.014

Rogerson, T., Cai, D. J., Frank, A., Sano, Y., Shobe, J., Lopez-Aranda, M. F., et al. (2014). Synaptic tagging during memory allocation. *Nat. Rev. Neurosci.* 15, 157–169. doi: 10.1038/nrn3667

Romano, S. A., Pérez-Schuster, V., Jouary, A., Boulanger-Weill, J., Candeo, A., Pietri, T., et al. (2017). An integrated calcium imaging processing toolbox for the analysis of neuronal population dynamics. *PLoS Comput. Biol.* 13:e1005526. doi: 10.1371/journal.pcbi.1005526

Soltys, Z., Orzylowska-Sliwinska, O., Zaremba, M., Orlowski, D., Piechota, M., Fiedorowicz, A., et al. (2005). Quantitative morphological study of microglial cells in the ischemic rat brain using principal component analysis. *J. Neurosci. Methods* 146, 50–60. doi: 10.1016/j.jneumeth.2005.01.009

Verhoeven, K. J., Simonsen, K. L., and Mcintyre, L. M. (2005). Implementing false discovery rate control: increasing your power. *Oikos* 108, 643–647. doi: 10.1111/j.0030-1299.2005.13727.x

Vousden, D. A., Epp, J., Okuno, H., Nieman, B. J., Van Eede, M., Dazai, J., et al. (2015). Whole-brain mapping of behaviourally induced neural activation in mice. *Brain Struct. Funct.* 220, 2043–2057. doi: 10.1007/s00429-014-0774-0

Walton, W. H. (1948). Ferets statistical diameter as a measure of particle size. *Nature* 162, 329–330. doi: 10.1038/162329b0

Zuur, A., Ieno, E. N., and Smith, G. M. (2007). *Analyzing Ecological Data*. Berlin: Springer Science & Business Media. doi: 10.1007/978-0-387-45972-1

Zuur, A. F., and Ieno, E. N. (2016). A protocol for conducting and presenting results of regression-type analyses. *Methods Ecol. Evol.* 7, 636–645. doi: 10.1111/2041-210X.12577

Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A. A., and Smith, G. M. (2009). "Zero-truncated and zero-inflated models for count data," in *Mixed Effects Models and Extensions in Ecology with R*, eds M. Gail, K. Krickeberg, J. M. Samet, A. Tsiatis, and W. Wong (Berlin: Springer), 261–293. doi: 10.1007/978-0-387-87458-6_11

frontiers in Neural Circuits

Auditory and Visual Motion Processing and Integration in the Primate Cerebral Cortex

Tristan A. Chaplin[1,2]*, Marcello G. P. Rosa[1,2] and Leo L. Lui[1,2]*

[1]Neuroscience Program, Biomedicine Discovery Institute and Department of Physiology, Monash University, Clayton, VIC, Australia, [2]Australian Research Council (ARC) Centre of Excellence for Integrative Brain Function, Monash University Node, Clayton, VIC, Australia

The ability of animals to detect motion is critical for survival, and errors or even delays in motion perception may prove costly. In the natural world, moving objects in the visual field often produce concurrent sounds. Thus, it can highly advantageous to detect motion elicited from sensory signals of either modality, and to integrate them to produce more reliable motion perception. A great deal of progress has been made in understanding how visual motion perception is governed by the activity of single neurons in the primate cerebral cortex, but far less progress has been made in understanding both auditory motion and audiovisual motion integration. Here we, review the key cortical regions for motion processing, focussing on translational motion. We compare the representations of space and motion in the visual and auditory systems, and examine how single neurons in these two sensory systems encode the direction of motion. We also discuss the way in which humans integrate of audio and visual motion cues, and the regions of the cortex that may mediate this process.

Keywords: visual motion, auditory motion, audiovisual integration, primates, cerebral cortex

The natural world abounds with motion, making this a highly salient cue to guide animals in interacting with the environment. It is therefore not surprising that most, if not all brains have dedicated neural circuits for the perception of motion. In primates, the cerebral cortex contains a network of regions that are specialized for motion processing, but the systems for processing the motion of visual features and sounds are mediated by different brain regions, and underpinned by different physiological mechanisms. In this mini-review article, we will discuss the encoding of direction of motion in the visual and auditory systems, with emphasis on the cortical systems that are involved in translational motion, especially in azimuth (leftwards and rightwards motion), as this is the most common type of motion used in audiovisual integration studies.

ENCODING OF DIRECTION OF MOTION IN THE ACTIVITY OF CORTICAL NEURONS

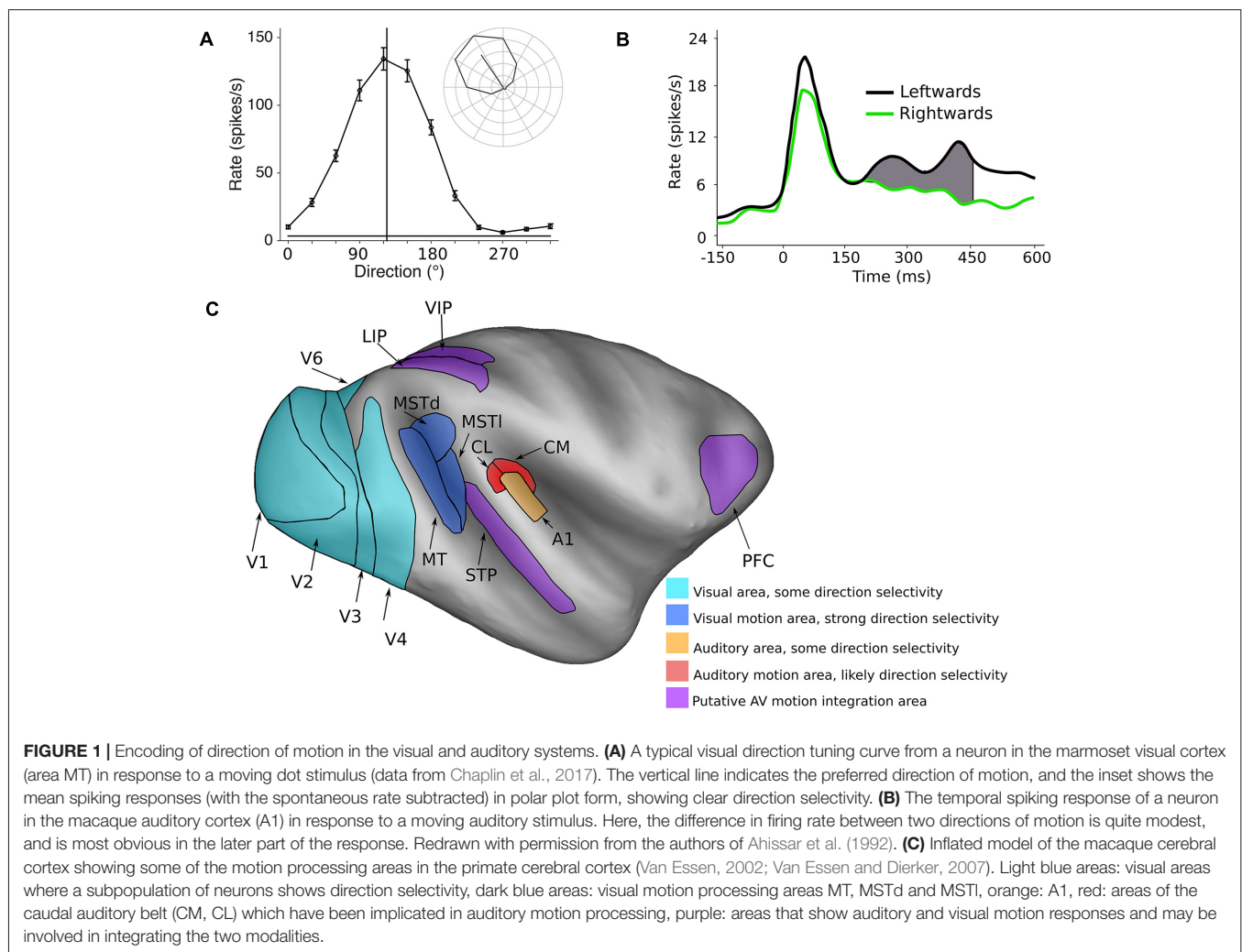Spatial features are represented in fundamentally different ways in the visual and auditory systems. In the visual system, most neurons have spatially defined receptive fields, which are ultimately defined by inputs from specific regions of the retina. Therefore, the responses of neurons in the visual system are inherently capable of coding the spatial location of visual stimuli, and in theory,

could encode direction of motion by the sequential activation of populations of neurons with different receptive field locations. However, the visual system goes one step further, with direction of motion being explicitly represented at the level of the single cell. Specifically, the spiking (action potential) responses of neurons are tuned to the direction of moving stimuli, meaning that they are more active in response to a specific direction of motion compared to other directions (Dubner and Zeki, 1971; Baker et al., 1981; Maunsell and Van Essen, 1983a; Albright, 1984; Desimone and Ungerleider, 1986; Saito et al., 1986; Tanaka and Saito, 1989; Chaplin et al., 2017). Thus, direction selective neurons in the visual system can encode the direction of motion *within* their receptive fields. For example, **Figure 1A** shows the response of a direction tuned neuron: the neuron shows strong responses to motion towards the upper left quadrant, and progressively weaker responses for directions further away.

In contrast, most neurons in the auditory system respond to specific ranges of acoustic frequencies, since they ultimately receive inputs from defined regions of the cochlea. Thus, the auditory system needs to exploit other auditory cues to extract spatial information from the stimulus. The principal

cues for locating sounds in the azimuth are binaural—interaural time differences (ITDs) and interaural level differences (ILDs; Middlebrooks and Green, 1991). Several brain regions are involved in the perception of sound location, and neurons in these regions can be tuned for ITDs or ILDs (Masterton et al., 1967; Rajan et al., 1990a,b; Semple and Kitzes, 1993a,b; Irvine et al., 1996; Tian et al., 2001; Woods et al., 2006; Miller and Recanzone, 2009; Grothe et al., 2010; Slee and Young, 2010; Kusmierek and Rauschecker, 2014; Keating and King, 2015; Lui et al., 2015; Mokri et al., 2015).

The encoding of the direction of auditory motion by the activity of single cortical neurons has not been studied extensively in primates—to our knowledge, there is only published study (Ahissar et al., 1992), in which they recorded spiking activity in the primary auditory cortex (A1) of monkeys. They found that while many cells (62%) in A1 code for the spatial location of stationary sounds, some cells (32%) also showed a preference for leftwards or rightwards direction of motion. However, the differences in responses were far less marked than those observed in direction selective cells in the visual system. There were only modest differences in firing rates, which were

FIGURE 1 | Encoding of direction of motion in the visual and auditory systems. **(A)** A typical visual direction tuning curve from a neuron in the marmoset visual cortex (area MT) in response to a moving dot stimulus (data from Chaplin et al., 2017). The vertical line indicates the preferred direction of motion, and the inset shows the mean spiking responses (with the spontaneous rate subtracted) in polar plot form, showing clear direction selectivity. **(B)** The temporal spiking response of a neuron in the macaque auditory cortex (A1) in response to a moving auditory stimulus. Here, the difference in firing rate between two directions of motion is quite modest, and is most obvious in the later part of the response. Redrawn with permission from the authors of Ahissar et al. (1992). **(C)** Inflated model of the macaque cerebral cortex showing some of the motion processing areas in the primate cerebral cortex (Van Essen, 2002; Van Essen and Dierker, 2007). Light blue areas: visual areas where a subpopulation of neurons shows direction selectivity, dark blue areas: visual motion processing areas MT, MSTd and MSTl, orange: A1, red: areas of the caudal auditory belt (CM, CL) which have been implicated in auditory motion processing, purple: areas that show auditory and visual motion responses and may be involved in integrating the two modalities.

evident in the late part of the responses (**Figure 1B**). These results suggest that the encoding of the direction of motion of auditory stimuli is likely to be a much more distributed representation across a neuronal populations, compared to direction of motion encoding in the visual system (Cohen and Newsome, 2009), or that explicit encoding of auditory motion relies on other areas beyond A1.

VISUAL MOTION PROCESSING AREAS

The neural circuits for visual motion processing are among the best understood aspects of the structure and function of the primate cerebral cortex (**Figure 1C**, blue areas). The primary visual cortex (V1) is the first stage of visual processing in the cerebral cortex in which direction selectivity first appears, but only a small proportion of V1 neurons are direction selective (∼15%, Yu et al., 2010; Yu and Rosa, 2014; Davies et al., 2016). Direction selective neurons have been observed in several other visual areas (Orban et al., 1986; Desimone and Schein, 1987; Felleman and Van Essen, 1987; Lui et al., 2005, 2006; Orban, 2008; Fattori et al., 2009; Li et al., 2013), but it is the middle temporal (MT) and medial superior temporal (MST) areas that appear to be most specialized for motion processing. The vast majority of cells in these regions are direction selective (MT ∼85%: Allman and Kaas, 1971; Dubner and Zeki, 1971; Maunsell and Van Essen, 1983b; Albright, 1984; MST ∼90%: Desimone and Ungerleider, 1986; Saito et al., 1986; Tanaka and Saito, 1989; Celebrini and Newsome, 1994; Elston and Rosa, 1997). Furthermore, it is known that damage to MT and MST results in motion perception impairments (Newsome and Paré, 1988; Pasternak and Merigan, 1994; Orban et al., 1995; Schenk and Zihl, 1997; Rudolph and Pasternak, 1999), and electrical stimulation of these regions can influence the perception of motion (Celebrini and Newsome, 1994, 1995; Salzman and Newsome, 1994; Britten and Van Wezel, 2002; Nichols and Newsome, 2002; Fetsch et al., 2014). Thus, a causal relationship has been established between neural activity in MT and MST and the perception of visual motion.

MST can be divided to two subregions: a lateral part (MSTl) involved in the perception of moving objects and smooth pursuit eye movements (Komatsu and Wurtz, 1988a,b; Eifuku and Wurtz, 1998), and dorsal part (MSTd), which is associated with the perception of complex motion patterns (Graziano et al., 1994; Mineault et al., 2012), especially self-motion (Saito et al., 1986; Komatsu and Wurtz, 1988a; Duffy and Wurtz, 1991; Duffy, 1998), and has a well described role in the integration of visual and vestibular motion cues (Gu et al., 2007, 2008). Differences between MT and MST have been well studied in monkeys, but in human studies these areas are typically grouped into a single region called the human MT complex (hMT+, Zeki et al., 1991; Huk et al., 2002), due to the spatial resolution limits of fMRI.

AUDITORY MOTION PROCESSING AREAS

In comparison to the visual system, the regions and circuitry of the cortex involved in auditory motion processing are not as well characterized (**Figure 1C**). While there is some evidence for

motion sensitivity and direction selectivity in the A1 (Ahissar et al., 1992; Griffiths et al., 2000; Lewis et al., 2000), many human imaging studies have identified the planum temporale, a region of auditory cortex caudal to primary cortex, as being the key site for auditory motion processing (Baumgart et al., 1999; Pavani et al., 2002; Warren et al., 2002; Alink et al., 2012b). In agreement with these findings, a recent imaging study in macaques also found that the caudal regions of auditory cortex are differentially activated by auditory motion compared to stationary stimuli (Poirier et al., 2017). Furthermore, studies of humans with lesions to caudal auditory cortex have found deficits in auditory motion processing (Ducommun et al., 2004; Lewald et al., 2009; Thaler et al., 2016).

It remains controversial whether auditory motion perception relies on specialized motion detectors, similar to direction selective cells in the visual cortex (Perrott and Musicant, 1977), or utilizes "snapshots" of the current sound source location (Ahissar et al., 1992; Poirier et al., 2017), as several human imaging studies have reported there is no difference in cortical activation between stationary and moving stimuli (Smith et al., 2004, 2007; Krumbholz et al., 2005, 2007). Since neurons in the auditory system show sensitivity to localization cues (e.g., ITDs and ILDs), the perception of motion could be mediated by the sequential activation of neurons that code for adjacent spatial locations (Ahissar et al., 1992). In general, in the auditory system the integration of binaural cues for sound localization occurs at early subcortical stages of processing, such as the superior olivary complex, the nuclei of the lateral lemniscus and the inferior colliculus (Moore, 1991). In monkeys, the caudal part of auditory cortex encompasses the caudomedial (CM) and caudolateral (CL) areas of the auditory belt (Hackett et al., 1998; Kaas et al., 1999), and these are known to play a role in the localization of auditory stimuli (Recanzone et al., 2000; Tian et al., 2001; Woods et al., 2006; Miller and Recanzone, 2009; Kusmierek and Rauschecker, 2014). Therefore, the sensitivity of neurons in these areas to the location of static stimuli is a potential confound in auditory motion studies, as it can be difficult to distinguish true motion sensitivity from sensitivity to spatial location. For example, it has been suggested that apparent sensitivities to motion in the inferior colliculus could be explained by adaptation to stationary stimuli, which would result in reduced spiking activity for stationary stimuli compared to moving stimuli (Ahissar et al., 1992; Wilson and O'Neill, 1998; McAlpine et al., 2000; Ingham et al., 2001; Poirier et al., 2017). However, the recent imaging study by Poirier et al. (2017) did take steps to control for this effect in their choice of stimuli and regressions analyses, and still found that the caudal auditory cortex was differentially activated by auditory motion compared to static motion. Further electrophysiological studies in monkeys will be required to address the question of how auditory motion is encoded by the spiking activity of neurons in these regions.

The neural representation of auditory motion does not necessarily have to be located in purely auditory regions. Direct reciprocal connections between MT/MST and the auditory cortex have been identified in primates (Palmer and Rosa, 2006), and two recent electrophysiological studies (Chaplin et al., 2018;

Kafaligonul et al., 2018) have reported evoked potentials in areas MT/MST in response to stationary auditory clicks. Two human imaging studies have reported that the hMT+ complex responds to auditory motion (Poirier et al., 2005; Strnad et al., 2013), but it has also been argued that observed auditory responses in hMT+ could be explained by localization errors (Jiang et al., 2014), and no study has found any evidence for spiking activity in response to auditory stimuli (moving or stationary) in the monkey MT complex. Furthermore, a case study of involving lesions of hMT+ did not find any impairment in the perception of auditory motion (Zihl et al., 1983). Thus, current evidence suggests that MT and MST are not involved in auditory motion processing.

INTEGRATION OF AUDITORY AND VISUAL MOTION CUES

Given the differences in the neural representation of motion in the auditory and visual systems, it is interesting to consider how the information from the two modalities could be combined to improve motion perception. Psychophysical studies have investigated audiovisual motion integration in humans using motion detection tasks, and have provided valuable insights into how auditory and visual motion can be integrated in the brain. Some of these studies have reported that humans perform better in audiovisual motion tasks compared to unimodal tasks, but there is disagreement as to whether this increase in performance is "statistically optimal" or the result of "probability summation." When probability summation occurs, observers perform better on bimodal trials because they essentially have two chances to answer correctly—using either the visual or the auditory cue (Wuerger et al., 2003; Alais and Burr, 2004). When statistically optimal integration occurs, observers combine the information obtained by the different senses by weighting according to their reliability, to make optimal use of the information available (Meyer and Wuerger, 2001). Therefore, statistically optimal integration exceeds the performance of probability summation. Multisensory integration has shown be statistically optimal in other contexts (Ernst and Banks, 2002; Angelaki et al., 2009; Fetsch et al., 2009; Drugowitsch et al., 2014; Rohde et al., 2016).

It has been argued that statistically optimal integration of multisensory cues relies on neural computations occurring in early sensory cortex (e.g., MT/MST), rather than in higher-level areas (Ma et al., 2006; Beck et al., 2008; Bizley et al., 2016). In contrast, when multisensory integration is the result of probability summation, it may rely on higher-order areas (e.g., prefrontal or posterior parietal cortex, Alais and Burr, 2004; Bizley et al., 2016).

AUDIOVISUAL MOTION INTEGRATION IN THE PRIMATE CEREBRAL CORTEX

Human imaging studies and monkey electrophysiological/ anatomical studies have suggested several candidate cortical regions for the integration of audiovisual motion. The human superior temporal sulcus is typically activated by moving audiovisual stimuli (Lewis et al., 2000; Baumann and Greenlee, 2007; von Saldern and Noppeney, 2013). This region likely corresponds to the superior temporal polysensory (STP) area of macaques (Bruce et al., 1981), and the presence of multisensory neurons in STP is well known (Bruce et al., 1981; Hikosaka et al., 1988; Watanabe and Iwai, 1991). STP is typically associated with processing more complex visual and auditory signals, such as faces and speech (Beauchamp, 2005) and biological motion (Oram and Perrett, 1994; Barraclough et al., 2005), especially in complex tasks (Meyer et al., 2011; Wuerger et al., 2012), but there is evidence of subregional specializations (Padberg et al., 2003).

The posterior parietal cortex may also be important for audiovisual motion integration, as areas in this complex have found to be active during audiovisual stimulation in humans (Baumann and Greenlee, 2007; Wuerger et al., 2012), and is thought play a key role in coordinating multisensory integration (Brang et al., 2013). Cells in the ventral intraparietal area (VIP) are known to respond to both visual motion (Cook and Maunsell, 2002; Kaminiarz et al., 2014) and auditory stimuli (Bremmer et al., 2001; Schlack et al., 2005). The lateral intraparietal area (LIP) has been demonstrated to be involved in the integration visual motion signals over time to form perceptual decisions (Roitman and Shadlen, 2002), and also responds to auditory stimulation (Grunewald et al., 1999; Linden et al., 1999). Therefore, it is possible that LIP could integration information from both senses, by preforming similar computations.

Integration could also occur at the level of the prefrontal cortex (PFC), as regions in the dorsolateral PFC (areas 8a, 45 and 46) are known to receive inputs from MT and MST (Lewis and Van Essen, 2000; Reser et al., 2013) as well as caudal auditory cortex (Romanski et al., 1999a,b). Furthermore, direction selective responses to visual motion have been demonstrated in this region (Zaksas and Pasternak, 2006), and like LIP, PFC neurons show activity that is consistent with accumulating sensory evidence to form perceptual decisions (Kim and Shadlen, 1999). Cells in the ventrolateral subdivision of the PFC, such as area 12, have been shown to integrate audiovisual cues, but like STP, are generally associated with higher level sensory processing, responding to individual faces and calls (Romanski, 2007, 2012). However, human imaging studies of audiovisual motion have generally not reported comparable activation in the PFC (Lewis et al., 2000; Baumann and Greenlee, 2007; von Saldern and Noppeney, 2013), although audiovisual biological motion can modulate activity in premotor areas (areas 6R and 44) when there is a mismatch between the auditory and visual cues (Wuerger et al., 2012).

A number of imaging studies have also found that audiovisual stimulation produces distinct activation (compared to visual only stimulation) in hMT+ (Alink et al., 2008; Lewis and Noppeney, 2010; Strnad et al., 2013; von Saldern and Noppeney, 2013), suggesting that auditory stimuli can modulate visually evoked responses (although this is not always the case, e.g., Wuerger et al., 2012). These regions receive sparse inputs from auditory cortex (Palmer and Rosa, 2006), and show evoked potentials in response to auditory stimuli (Chaplin et al., 2018;

Kafaligonul et al., 2018). Additionally, auditory motion has been shown to affect various aspects of visual perception, such as improving visual motion detection (Kim et al., 2012), improve learning in visual motion tasks (Seitz et al., 2006), and induce visual illusions (Sekuler et al., 1997; Meyer and Wuerger, 2001; Kitagawa and Ichihara, 2002; Beer and Röder, 2004; Soto-Faraco et al., 2005; Freeman and Driver, 2008; Alink et al., 2012a; Kafaligonul and Stoner, 2012; Kafaligonul and Oluk, 2015). Altogether, these studies suggest that auditory stimuli, especially when moving, could modulate responses to visual stimuli in MT/MST.

To specifically test this hypothesis, we have investigated if auditory motion cues are integrated with visual motion cues in MT/MST, by recording spiking activity and characterizing the ability of neurons to encode the direction of motion, using ideal observer analysis (Chaplin et al., 2018). We presented random dot patterns that moved either leftwards or rightwards, and manipulated the strength of the visual motion signal by reducing the coherence of the dots (i.e., making some proportion of the dots move in random directions). Reducing motion coherence reduces the both the psychophysical performance of observers (i.e., makes it more difficult to discriminate the directions of motion) and the neurometric performance of single neurons (i.e., reduces the neuronal information; Newsome et al., 1989). We hypothesized that the addition of an auditory stimulus that moved in the same direction as the visual stimulus would increase the information carried by single neurons and therefore increase neurometric performance, just as it can increase psychophysical performance in humans (Meyer and Wuerger, 2001; Kim et al., 2012). In particular, we predicted that auditory cues would be most likely be integrated at low motion coherence levels, in line with Bayesian models of multisensory integration (Ernst and Banks, 2002; Ma et al., 2006; Gu et al., 2008). However, we found no evidence of spike rate modulations (**Figure 2A**) or improvements in neurometric performance (**Figure 2B**) due

FIGURE 2 | (A) Responses of a marmoset MT neuron to visual, auditory and audiovisual stimuli. The raster plots (black dots) and spike rate functions (colored lines) show a clear response to visual but not auditory stimuli (blue vs. green lines). The combination of auditory and visual stimuli (red line) was not significantly different to the visual only response (blue vs. red lines). **(B)** Neurometric performance (measured as the area under the receiver operating characteristic (ROC) curve, Britten et al., 1992, which corresponds to the performance of an ideal observer discriminating the direction of motion using the spiking activity of the neuron) of a marmoset MT neuron when discriminating leftwards and rightwards motion under visual (blue) and audiovisual (red) conditions at different levels of motion coherence (strength of motion signal). The addition of the auditory stimulus did not shift the neurometric curve to the left as would be expected if the neuron was integrating the auditory motion cue (adapted from Chaplin et al., 2018).

to the auditory stimulus, in MT or MST. It may be the case that the audiovisual responses observed in hMT+ are the result of task related signals (Alink et al., 2012b; Bizley et al., 2016; Kayser et al., 2017), such as the binding of the two modalities to form a unified percept (Nahorna et al., 2012, 2015; Bizley and Cohen, 2013), attentional effects (Beer and Röder, 2004, 2005; Lakatos et al., 2008), or choice-related signals from the decision making process (Cumming and Nienborg, 2016).

Only one other study has investigated the effects of auditory stimuli on the responses of MT neurons (Kafaligonul et al., 2018). This study aimed to test if the activity of MT neurons mediated the temporal ventriloquist illusion, in which stationary auditory clicks induce influence the perception of visual speed. The authors hypothesized that the auditory clicks would alter the speed tuning and response duration of MT neurons in response to apparent visual motion. However, the auditory stimuli did not alter speed tuning or response duration in a way that would support the perception of the illusion, even though there was a possible modulation of the temporal spiking response. Therefore, electrophysiological studies in monkeys so far suggest that auditory stimuli do not influence visual motion perception through changes in activity to MT/MST neurons. However, since the projections from auditory to visual cortex are known to arrive at the peripheral representation of the visual field (Palmer and Rosa, 2006; Majka et al., 2018), it possible that their role of auditory inputs to facilitate the detection and localization of visual features, especially for orienting (Perrott et al., 1993; Wang et al., 2008).

CONCLUSION

In conclusion, the processing of auditory and visual motion in the primate cerebral cortex utilizes different brain areas and physiological mechanisms. While good progress has been made in identifying the cortical regions involved in processing auditory and audiovisual motion, the mechanisms of audiovisual integration remain unclear. The current evidence from single neuron studies suggests that the integration of auditory and visual motion cues is not mediated by the early visual areas MT and MST, and therefore such integration likely occurs in higher level cortical areas. Another possibility is that the integration of audiovisual motion signals is not mediated by a single brain region, but instead by synchronized network activity (Lewis and Noppeney, 2010).

AUTHOR CONTRIBUTIONS

TC wrote the first draft of the manuscript. MR and LL wrote sections of the manuscript. All authors contributed to manuscript revision, read and approved the submitted version.

FUNDING
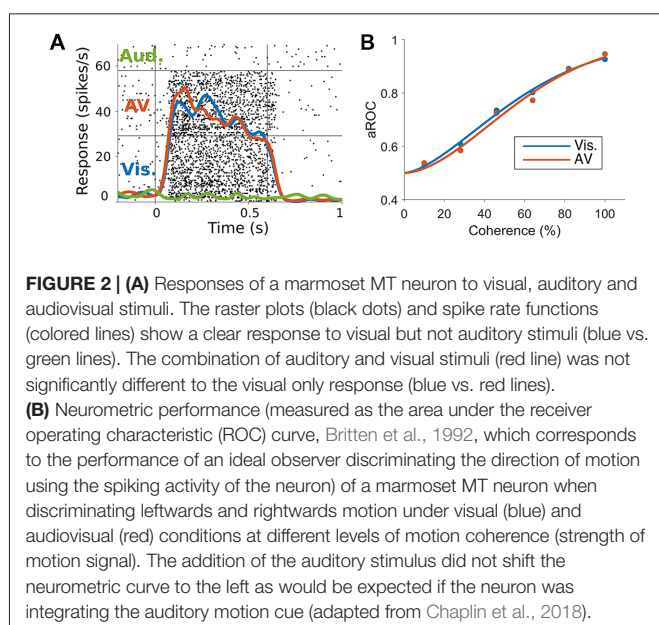
REFERENCES

Ahissar, M., Ahissar, E., Bergman, H., and Vaadia, E. (1992). Encoding of sound-source location and movement: activity of single neurons and interactions between adjacent neurons in the monkey auditory cortex. *J. Neurophysiol.* 67, 203–215. doi: 10.1152/jn.1992.67.1.203

Alais, D., and Burr, D. C. (2004). No direction-specific bimodal facilitation for audiovisual motion detection. *Cogn. Brain Res.* 19, 185–194. doi: 10.1016/j.cogbrainres.2003.11.011

Albright, T. D. (1984). Direction and orientation selectivity of neurons in visual area MT of the macaque. *J. Neurophysiol.* 52, 1106–1130. doi: 10.1152/jn.1984.52.6.1106

Alink, A., Euler, F., Galeano, E., Krugliak, A., Singer, W., and Kohler, A. (2012a). Auditory motion capturing ambiguous visual motion. *Front. Psychol.* 3, 1–8. doi: 10.3389/fpsyg.2011.00391

Alink, A., Euler, F., Kriegeskorte, N., Singer, W., and Kohler, A. (2012b). Auditory motion direction encoding in auditory cortex and high-level visual cortex. *Hum. Brain Mapp.* 33, 969–978. doi: 10.1002/hbm.21263

Alink, A., Singer, W., and Muckli, L. (2008). Capture of auditory motion by vision is represented by an activation shift from auditory to visual motion cortex. *J. Neurosci.* 28, 2690–2697. doi: 10.1523/jneurosci.2980-07.2008

Allman, J. M., and Kaas, J. H. (1971). A representation of the visual field in the caudal third of the middle temporal gyrus of the owl monkey (*Aotus trivirgatus*). *Brain Res.* 31, 85–105. doi: 10.1016/0006-8993(71)90635-4

Angelaki, D. E., Gu, Y., and DeAngelis, G. C. (2009). Multisensory integration: psychophysics, neurophysiology and computation. *Curr. Opin. Neurobiol.* 19, 452–458. doi: 10.1016/j.conb.2009.06.008

Baker, J. F., Petersen, S. E., Newsome, W. T., and Allman, J. M. (1981). Visual response properties of neurons in four extrastriate visual areas of the owl monkey (Aotus trivirgatus): a quantitative comparison of medial, dorsomedial, dorsolateral and middle temporal areas. *J. Neurophysiol.* 45, 397–416. doi: 10.1152/jn.1981.45.3.397

Barraclough, N. E., Xiao, D., Baker, C. I., Oram, M. W., and Perrett, D. I. (2005). Integration of visual and auditory information by superior temporal sulcus neurons responsive to the sight of actions. *J. Cogn. Neurosci.* 17, 377–391. doi: 10.1162/0898929053279586

Baumann, O., and Greenlee, M. W. (2007). Neural correlates of coherent audiovisual motion perception. *Cereb. Cortex* 17, 1433–1443. doi: 10.1093/cercor/bhl055

Baumgart, F., Gaschler-Markefski, B., Woldorff, M. G., Heinze, H. J., and Scheich, H. (1999). A movement-sensitive area in auditory cortex. *Nature* 400, 724–726. doi: 10.1038/23385

Beauchamp, M. S. (2005). See me, hear me, touch me: multisensory integration in lateral occipital-temporal cortex. *Curr. Opin. Neurobiol.* 15, 145–153. doi: 10.1016/j.conb.2005.03.011

Beck, J. M., Ma, W. J., Kiani, R., Hanks, T., Churchland, A. K., Roitman, J. D., et al. (2008). Probabilistic population codes for Bayesian decision making. *Neuron* 60, 1142–1152. doi: 10.1016/j.neuron.2008.09.021

Beer, A. L., and Röder, B. (2004). Unimodal and crossmodal effects of endogenous attention to visual and auditory motion. *Cogn. Affect. Behav. Neurosci.* 4, 230–240. doi: 10.3758/cabn.4.2.230

Beer, A. L., and Röder, B. (2005). Attending to visual or auditory motion affects perception within and across modalities: an event-related potential study. *Eur. J. Neurosci.* 21, 1116–1130. doi: 10.1111/j.1460-9568.2005.03927.x

Bizley, J. K., and Cohen, Y. E. (2013). The what, where and how of auditory-object perception. *Nat. Rev. Neurosci.* 14, 693–707. doi: 10.1038/nrn3565

Bizley, J. K., Jones, G. P., and Town, S. M. (2016). Where are multisensory signals combined for perceptual decision-making? *Curr. Opin. Neurobiol.* 40, 31–37. doi: 10.1016/j.conb.2016.06.003

Brang, D., Taich, Z. J., Hillyard, S. A., Grabowecky, M., and Ramachandran, V. S. (2013). Parietal connectivity mediates multisensory facilitation. *Neuroimage* 78, 396–401. doi: 10.1016/j.neuroimage.2013.04.047

ACKNOWLEDGMENTS

Bremmer, F., Schlack, A., Shah, N. J., Zafiris, O., Kubischik, M., Hoffmann, K., et al. (2001). Polymodal motion processing in posterior parietal and premotor cortex: a human fMRI study strongly implies equivalencies between humans and monkeys. *Neuron* 29, 287–296. doi: 10.1016/s0896-6273(01)00198-2

Britten, K. H., Shadlen, M. N., Newsome, W. T., and Movshon, J. A. (1992). The analysis of visual motion: a comparison of neuronal and psychophysical performance. *J. Neurosci.* 12, 4745–4765. doi: 10.1523/jneurosci.12-12-04745.1992

Britten, K. H., and Van Wezel, R. J. A. (2002). Area MST and heading perception in macaque monkeys. *Cereb. Cortex* 12, 692–701. doi: 10.1093/cercor/12.7.692

Bruce, C., Desimone, R., and Gross, C. G. (1981). Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. *J. Neurophysiol.* 46, 369–384. doi: 10.1152/jn.1981.46.2.369

Celebrini, S., and Newsome, W. T. (1994). Neuronal and psychophysical sensitivity to motion signals in extrastriate area MST of the macaque monkey. *J. Neurosci.* 14, 4109–4124. doi: 10.1523/jneurosci.14-07-04109.1994

Celebrini, S., and Newsome, W. T. (1995). Microstimulation of extrastriate area MST influences performance on a direction discrimination task. *J. Neurophysiol.* 73, 437–448. doi: 10.1152/jn.1995.73.2.437

Chaplin, T. A., Allitt, B. J., Hagan, M. A., Price, N. S. C., Rajan, R., Rosa, M. G. P., et al. (2017). Sensitivity of neurons in the middle temporal area of marmoset monkeys to random dot motion. *J. Neurophysiol.* 118, 1567–1580. doi: 10.1101/104117

Chaplin, T. A., Allitt, B. J., Hagan, M. A., Rosa, M. G. P., Rajan, R., and Lui, L. L. (2018). Auditory motion does not modulate spiking activity in the middle temporal and medial superior temporal visual areas. *Eur. J. Neurosci.* 48, 2013–2029. doi: 10.1111/ejn.14071

Cohen, M. R., and Newsome, W. T. (2009). Estimates of the contribution of single neurons to perception depend on timescale and noise correlation. *J. Neurosci.* 29, 6635–6648. doi: 10.1523/jneurosci.5179-08.2009

Cook, E. P., and Maunsell, J. H. R. (2002). Dynamics of neuronal responses in macaque MT and VIP during motion detection. *Nat. Neurosci.* 5, 985–994. doi: 10.1038/nn924

Cumming, B. G., and Nienborg, H. (2016). Feedforward and feedback sources of choice probability in neural population responses. *Curr. Opin. Neurobiol.* 37, 126–132. doi: 10.1016/j.conb.2016.01.009

Davies, A. J., Chaplin, T. A., Rosa, M. G. P., and Yu, H.-H. (2016). Natural motion trajectory enhances the coding of speed in primate extrastriate cortex. *Sci. Rep.* 6:19739. doi: 10.1038/srep19739

Desimone, R., and Schein, S. J. (1987). Visual properties of neurons in area V4 of the macaque: sensitivity to stimulus form. *J. Neurophysiol.* 57, 835–868. doi: 10.1152/jn.1987.57.3.835

Desimone, R., and Ungerleider, L. G. (1986). Multiple visual areas in the caudal superior temporal sulcus of the macaque. *J. Comp. Neurol.* 248, 164–189. doi: 10.1002/cne.902480203

Drugowitsch, J., DeAngelis, G. C., Klier, E. M., Angelaki, D. E., and Pouget, A. (2014). Optimal multisensory decision-making in a reaction-time task. *Elife* 3:e03005. doi: 10.7554/eLife.03005

Dubner, R., and Zeki, S. M. (1971). Response properties and receptive fields of cells in an anatomically defined region of the superior temporal sulcus in the monkey. *Brain Res.* 35, 528–532. doi: 10.1016/0006-8993(71)90494-x

Ducommun, C. Y., Michel, C. M., and Clarke, S. (2004). Cortical motion deafness. *Neuron* 43, 765–777. doi: 10.1016/j.neuron.2004.08.020

Duffy, C. J. (1998). MST neurons respond to optic flow and translational movement. *J. Neurophysiol.* 80, 1816–1827. doi: 10.1152/jn.1998.80.4.1816

Duffy, C. J., and Wurtz, R. H. (1991). Sensitivity of MST neurons to optic flow stimuli. II. Mechanisms of response selectivity revealed by small-field stimuli. *J. Neurophysiol.* 65, 1346–1359. doi: 10.1152/jn.1991.65.6.1346

Eifuku, S., and Wurtz, R. H. (1998). Response to motion in extrastriate area MSTl: center-surround interactions. *J. Neurophysiol.* 80, 282–296. doi: 10.1152/jn.1998.80.1.282

Elston, G. N., and Rosa, M. G. (1997). The occipitoparietal pathway of the macaque monkey: comparison of pyramidal cell morphology in layer III of functionally related cortical visual areas. *Cereb. Cortex* 7, 432–452. doi: 10.1093/cercor/7. 5.432

Ernst, M. O., and Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* 415, 429–433. doi: 10.1038/415429a

Fattori, P., Pitzalis, S., and Galletti, C. (2009). The cortical visual area V6 in macaque and human brains. *J. Physiol. Paris* 103, 88–97. doi: 10.1016/j. jphysparis.2009.05.012

Felleman, D. J., and Van Essen, D. C. (1987). Receptive field properties of neurons in area V3 of macaque monkey extrastriate cortex. *J. Neurophysiol.* 57, 889–920. doi: 10.1152/jn.1987.57.4.889

Fetsch, C. R., Kiani, R., Newsome, W. T., and Shadlen, M. N. (2014). Effects of cortical microstimulation on confidence in a perceptual decision. *Neuron* 83, 797–804. doi: 10.1016/j.neuron.2014.07.011

Fetsch, C. R., Turner, A. H., DeAngelis, G. C., and Angelaki, D. E. (2009). Dynamic reweighting of visual and vestibular cues during self-motion perception. *J. Neurosci.* 29, 15601–15612. doi: 10.1523/JNEUROSCI.2574-09.2009

Freeman, E., and Driver, J. (2008). Direction of visual apparent motion driven solely by timing of a static sound. *Curr. Biol.* 18, 1262–1266. doi: 10.1016/j.cub. 2008.07.066

Graziano, M. S., Andersen, R. A., and Snowden, R. J. (1994). Tuning of MST neurons to spiral motions. *J. Neurosci.* 14, 54–67. doi: 10.1523/JNEUROSCI. 14-01-00054.1994

Griffiths, T. D., Green, G. G., Rees, A., and Rees, G. (2000). Human brain areas involved in the analysis of auditory movement. *Hum. Brain Mapp.* 9, 72–80. doi: 10.1002/(sici)1097-0193(200002)9:2<72::aid-hbm2>3.0.co;2-9

Grothe, B., Pecka, M., and McAlpine, D. (2010). Mechanisms of sound localization in mammals. *Physiol. Rev.* 90, 983–1012. doi: 10.1152/physrev. 00026.2009

Grunewald, A., Linden, J. F., and Andersen, R. A. (1999). Responses to auditory stimuli in macaque lateral intraparietal area I. Effects of training. *J. Neurophysiol.* 82, 330–342. doi: 10.1152/jn.1999.82.1.330

Gu, Y., DeAngelis, G. C., and Angelaki, D. E. (2007). A functional link between area MSTd and heading perception based on vestibular signals. *Nat. Neurosci.* 10, 1038–1047. doi: 10.1038/nn1935

Gu, Y., Angelaki, D. E., and DeAngelis, G. C. (2008). Neural correlates of multisensory cue integration in macaque MSTd. *Nat. Neurosci.* 11, 1201–1210. doi: 10.1038/nn.2191

Hackett, T. A., Stepniewska, I., and Kaas, J. H. (1998). Subdivisions of auditory cortex and ipsilateral cortical connections of the parabelt auditory cortex in macaque monkeys. *J. Comp. Neurol.* 394, 475–495. doi: 10.1002/(sici)1096-9861(19980518)394:4<475::aid-cne6>3.0.co;2-z

Hikosaka, K., Iwai, E., Saito, H., and Tanaka, K. (1988). Polysensory properties of neurons in the anterior bank of the caudal superior temporal sulcus of the macaque monkey. *J. Neurophysiol.* 60, 1615–1637. doi: 10.1152/jn.1988.60. 5.1615

Huk, A. C., Dougherty, R. F., and Heeger, D. J. (2002). Retinotopy and functional subdivision of human areas MT and MST. *J. Neurosci.* 22, 7195–7205. doi: 10.1523/jneurosci.22-16-07195.2002

Ingham, N. J., Hart, H. C., and McAlpine, D. (2001). Spatial receptive fields of inferior colliculus neurons to auditory apparent motion in free field. *J. Neurophysiol.* 85, 23–33. doi: 10.1152/jn.2001.85.1.23

Irvine, D. R., Rajan, R., and Aitkin, L. M. (1996). Sensitivity to interaural intensity differences of neurons in primary auditory cortex of the cat. I. types of sensitivity and effects of variations in sound pressure level. *J. Neurophysiol.* 75, 75–96. doi: 10.1152/jn.1996.75.1.75

Jiang, F., Stecker, G. C., and Fine, I. (2014). Auditory motion processing after early blindness. *J. Vis.* 14:4. doi: 10.1167/14.13.4

Kaas, J. H., Hackett, T. A., and Tramo, M. J. (1999). Auditory processing in primate cerebral cortex. *Curr. Opin. Neurobiol.* 9, 164–170. doi: 10.1016/S0959-4388(99)80022-1

Kafaligonul, H., Albright, T. D., and Stoner, G. R. (2018). Auditory modulation of spiking activity and local field potentials in area MT does not appear to underlie an audiovisual temporal illusion. *J. Neurophysiol.* 120, 1340–1355. doi: 10.1152/jn.00835.2017

Kafaligonul, H., and Oluk, C. (2015). Audiovisual associations alter the perception of low-level visual motion. *Front. Integr. Neurosci.* 9:26. doi: 10.3389/fnint.2015. 00026

Kafaligonul, H., and Stoner, G. R. (2012). Static sound timing alters sensitivity to low-level visual motion. *J. Vis.* 12:2. doi: 10.1167/12.11.2

Kaminiarz, A., Schlack, A., Hoffmann, K.-P., Lappe, M., and Bremmer, F. (2014). Visual selectivity for heading in the macaque ventral intraparietal area. *J. Neurophysiol.* 112, 2470–2480. doi: 10.1152/jn.00410.2014

Kayser, S. J., Philiastides, M. G., and Kayser, C. (2017). Sounds facilitate visual motion discrimination via the enhancement of late occipital visual representations. *Neuroimage* 148, 31–41. doi: 10.1016/j.neuroimage.2017. 01.010

Keating, P., and King, A. J. (2015). Sound localization in a changing world. *Curr. Opin. Neurobiol.* 35, 35–43. doi: 10.1016/j.conb.2015.06.005

Kim, R., Peters, M. A., and Shams, L. (2012). 0 + 1 > 1: how adding noninformative sound improves performance on a visual task. *Psychol. Sci.* 23, 6–12. doi: 10.1177/0956797611420662

Kim, J. N., and Shadlen, M. N. (1999). Neural correlates of a decision in the dorsolateral prefrontal cortex of the macaque. *Nat. Neurosci.* 2, 176–185. doi: 10.1038/5739

Kitagawa, N., and Ichihara, S. (2002). Hearing visual motion in depth. *Nature* 416, 172–174. doi: 10.1038/416172a

Komatsu, H., and Wurtz, R. H. (1988a). Relation of cortical areas MT and MST to pursuit eye movements. I. Localization and visual properties of neurons. *J. Neurophysiol.* 60, 580–603. doi: 10.1152/jn.1988.60.2.580

Komatsu, H., and Wurtz, R. H. (1988b). Relation of cortical areas MT and MST to pursuit eye movements. III. Interaction with full-field visual stimulation. *J. Neurophysiol.* 60, 621–644. doi: 10.1152/jn.1988.60.2.621

Krumbholz, K., Hewson-Stoate, N., and Schönwiesner, M. (2007). Cortical response to auditory motion suggests an asymmetry in the reliance on inter-hemispheric connections between the left and right auditory cortices. *J. Neurophysiol.* 97, 1649–1655. doi: 10.1152/jn.00560.2006

Krumbholz, K., Schönwiesner, M., Rübsamen, R., Zilles, K., Fink, G. R., and Von Cramon, D. Y. (2005). Hierarchical processing of sound location and motion in the human brainstem and planum temporale. *Eur. J. Neurosci.* 21, 230–238. doi: 10.1111/j.1460-9568.2004.03836.x

Kusmierek, P., and Rauschecker, J. P. (2014). Selectivity for space and time in early areas of the auditory dorsal stream in the rhesus monkey. *J. Neurophysiol.* 111, 1671–1685. doi: 10.1152/jn.00436.2013

Lakatos, P., Karmos, G., Mehta, A. D., Ulbert, I., and Schroeder, C. E. (2008). Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science* 320, 110–113. doi: 10.1126/science.1154735

Lewald, J., Peters, S., Corballis, M. C., and Hausmann, M. (2009). Perception of stationary and moving sound following unilateral cortectomy. *Neuropsychologia* 47, 962–971. doi: 10.1016/j.neuropsychologia.2008. 10.016

Lewis, J. W., Beauchamp, M. S., and DeYoe, E. A. (2000). A comparison of visual and auditory motion processing in human cerebral cortex. *Cereb. Cortex* 10, 873–888. doi: 10.1093/cercor/10.9.873

Lewis, R., and Noppeney, U. (2010). Audiovisual synchrony improves motion discrimination via enhanced connectivity between early visual and auditory areas. *J. Neurosci.* 30, 12329–12339. doi: 10.1523/JNEUROSCI.5745-09.2010

Lewis, J. W., and Van Essen, D. C. (2000). Corticocortical connections of visual, sensorimotor and multimodal processing areas in the parietal lobe of the macaque monkey. *J. Comp. Neurol.* 428, 112–137. doi: 10.1002/1096-9861(20001204)428:1<112::aid-cne8>3.0.co;2-9

Li, P., Zhu, S., Chen, M., Han, C., Xu, H., Hu, J., et al. (2013). A motion direction preference map in monkey V4. *Neuron* 78, 376–388. doi: 10.1016/j.neuron. 2013.02.024

Linden, J. F., Grunewald, A., Andersen, R. A., and Modulation, I. I. B. (1999). Responses to auditory stimuli in macaque lateral intraparietal area II. Behavioral modulation. *J. Neurophysiol.* 82, 343–358. doi: 10.1152/jn.1999.82. 1.343

Lui, L. L., Bourne, J. A., and Rosa, M. G. P. (2005). Single-unit responses to kinetic stimuli in New World monkey area V2: physiological characteristics of cue-invariant neurones. *Exp. Brain Res.* 162, 100–108. doi: 10.1007/s00221-004-2113-9

Lui, L. L., Bourne, J. A., and Rosa, M. G. P. (2006). Functional response properties of neurons in the dorsomedial visual area of New World monkeys (*Callithrix jacchus*). *Cereb. Cortex* 16, 162–177. doi: 10.1093/cercor/bhi094

Lui, L. L., Mokri, Y., Reser, D. H., Rosa, M. G. P., and Rajan, R. (2015). Responses of neurons in the marmoset primary auditory cortex to interaural level differences: comparison of pure tones and vocalizations. *Front. Neurosci.* 9:132. doi: 10.3389/fnins.2015.00132

Ma, W. J., Beck, J. M., Latham, P. E., and Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nat. Neurosci.* 9, 1432–1438. doi: 10.1038/nn1790

Majka, P., Rosa, M. G. P., Bai, S., Chan, J. M., Huo, B.-X., Jermakow, N., et al. (2018). Unidirectional monosynaptic connections from auditory areas to the primary visual cortex in the marmoset monkey. *Brain Struct. Funct.* doi: 10.1007/s00429-018-1764-4 [Epub ahead of print].

Masterton, B., Diamond, I. T., Harrison, J. M., and Beecher, M. D. (1967). Medial superior olive and sound localization. *Science* 155, 1696–1697. doi: 10.1126/science.155.3770.1696-a

Maunsell, J. H. R., and Van Essen, D. C. (1983a). Functional properties of neurons in middle temporal visual area of the macaque monkey. I. Selectivity for stimulus direction, speed, and orientation. *J. Neurophysiol.* 49, 1127–1147. doi: 10.1152/jn.1983.49.5.1127

Maunsell, J. H. R., and Van Essen, D. C. (1983b). Functional properties of neurons in middle temporal visual area of the macaque monkey. II. Binocular interactions and sensitivity to binocular disparity. *J. Neurophysiol.* 49, 1148–1167. doi: 10.1152/jn.1983.49.5.1148

McAlpine, D., Jiang, D., Shackleton, T. M., and Palmer, A. R. (2000). Responses of neurons in the inferior colliculus to dynamic interaural phase cues: evidence for a mechanism of binaural adaptation. *J. Neurophysiol.* 83, 1356–1365. doi: 10.1152/jn.2000.83.3.1356

Meyer, G. F., Greenlee, M., and Wuerger, S. M. (2011). Interactions between auditory and visual semantic stimulus classes: evidence for common processing networks for speech and body actions. *J. Cogn. Neurosci.* 23, 2291–2308. doi: 10.1162/jocn.2010.21593

Meyer, G. F., and Wuerger, S. M. (2001). Cross-modal integration of auditory and visual motion signals. *Neuroreport* 12, 2557–2560. doi: 10.1097/00001756-200108080-00053

Middlebrooks, J., and Green, D. (1991). Sound localization by human listeners. *Annu. Rev. Psychol.* 42, 135–159. doi: 10.1146/annurev.psych.42.1.135

Miller, L. M., and Recanzone, G. H. (2009). Populations of auditory cortical neurons can accurately encode acoustic space across stimulus intensity. *Proc. Natl. Acad. Sci. U S A* 106, 5931–5935. doi: 10.1073/pnas.0901023106

Mineault, P. J., Khawaja, F. A., Butts, D. A., and Pack, C. C. (2012). Hierarchical processing of complex motion along the primate dorsal visual pathway. *Proc. Natl. Acad. Sci. U S A* 109, E972–E980. doi: 10.1073/pnas.1115685109

Mokri, Y., Worland, K., Ford, M., and Rajan, R. (2015). Effect of background noise on neuronal coding of interaural level difference cues in rat inferior colliculus. *Eur. J. Neurosci.* 42, 1685–1704. doi: 10.1111/ejn.12914

Moore, D. R. (1991). Anatomy and physiology of binaural hearing. *Audiology* 30, 125–134. doi: 10.3109/00206099109072878

Nahorna, O., Berthommier, F., and Schwartz, J.-L. (2012). Binding and unbinding the auditory and visual streams in the McGurk effect. *J. Acoust. Soc. Am.* 132, 1061–1077. doi: 10.1121/1.4728187

Nahorna, O., Berthommier, F., and Schwartz, J.-L. (2015). Audio-visual speech scene analysis: characterization of the dynamics of unbinding and rebinding the McGurk effect. *J. Acoust. Soc. Am.* 137, 362–377. doi: 10.1121/1.4904536

Newsome, W. T., Britten, K. H., and Movshon, J. A. (1989). Neuronal correlates of a perceptual decision. *Nature* 341, 52–54. doi: 10.1038/341052a0

Newsome, W. T., and Paré, E. B. (1988). A selective impairment of motion perception following lesions of the middle temporal visual area (MT). *J. Neurosci.* 8, 2201–2211. doi: 10.1523/jneurosci.08-06-02201.1988

Nichols, M. J., and Newsome, W. T. (2002). Middle temporal visual area microstimulation influences veridical judgments of motion direction. *J. Neurosci.* 22, 9530–9540. doi: 10.1523/jneurosci.22-21-09530.2002

Oram, M. W., and Perrett, D. I. (1994). Responses of anterior superior temporal polysensory (STPa) neurons to "biological motion" stimuli. *J. Cogn. Neurosci.* 6, 99–116. doi: 10.1162/jocn.1994.6.2.99

Orban, G. A. (2008). Higher order visual processing in macaque extrastriate cortex. *Physiol. Rev.* 88, 59–89. doi: 10.1152/physrev.00008.2007

Orban, G. A., Kennedy, H., and Bullier, J. (1986). Velocity sensitivity and direction selectivity of neurons in areas V1 and V2 of the monkey: influence of eccentricity. *J. Neurophysiol.* 56, 462–480. doi: 10.1152/jn.1986.56.2.462

Orban, G. A., Saunders, R. C., and Vandenbussche, E. (1995). Lesions of the superior temporal cortical motion areas impair speed discrimination in the macaque monkey. *Eur. J. Neurosci.* 7, 2261–2276. doi: 10.1111/j.1460-9568.1995.tb00647.x

Padberg, J., Seltzer, B., and Cusick, C. G. (2003). Architectonics and cortical connections of the upper bank of the superior temporal sulcus in the rhesus monkey: an analysis in the tangential plane. *J. Comp. Neurol.* 467, 418–434. doi: 10.1002/cne.10932

Palmer, S. M., and Rosa, M. G. P. (2006). A distinct anatomical network of cortical areas for analysis of motion in far peripheral vision. *Eur. J. Neurosci.* 24, 2389–2405. doi: 10.1111/j.1460-9568.2006.05113.x

Pasternak, T., and Merigan, W. H. (1994). Motion perception following lesions of the superior temporal sulcus in the monkey. *Cereb. Cortex* 4, 247–259. doi: 10.1093/cercor/4.3.247

Pavani, F., Macaluso, E., Warren, J., Driver, J., and Griffiths, T. (2002). A common cortical substrate activated by horizontal and vertical sound movement in the human brain. *Curr. Biol.* 12, 1584–1590. doi: 10.1016/s0960-9822(02)01143-0

Perrott, D. R., Costantino, B., and Cisneros, J. (1993). Auditory and visual localization performance in a sequential task. *J. Acoust. Soc. Am.* 93, 2134–2138. doi: 10.1121/1.406675

Perrott, D. R., and Musicant, A. D. (1977). Minimum auditory movement angle: binaural localization of moving sound sources. *J. Acoust. Soc. Am.* 62, 1463–1466. doi: 10.1121/1.381675

Poirier, C., Baumann, S., Dheerendra, P., Joly, O., Hunter, D., Balezeau, F., et al. (2017). Auditory motion-specific mechanisms in the primate brain. *PLoS Biol.* 15:e2001379. doi: 10.1371/journal.pbio.2001379

Poirier, C., Collignon, O., DeVolder, A. G., Renier, L., Vanlierde, A., Tranduy, D., et al. (2005). Specific activation of the V5 brain area by auditory motion processing: an fMRI study. *Cogn. Brain Res.* 25, 650–658. doi: 10.1016/j.cogbrainres.2005.08.015

Rajan, R., Aitkin, L. M., and Irvine, D. R. (1990a). Azimuthal sensitivity of neurons in primary auditory cortex of cats. II. Organization along frequency-band strips. *J. Neurophysiol.* 64, 888–902. doi: 10.1152/jn.1990.64.3.888

Rajan, R., Aitkin, L. M., Irvine, D. R., and McKay, J. (1990b). Azimuthal sensitivity of neurons in primary auditory cortex of cats. I. Types of sensitivity and the effects of variations in stimulus parameters. *J. Neurophysiol.* 64, 872–887. doi: 10.1152/jn.1990.64.3.872

Recanzone, G. H., Guard, D. C., Phan, M. L., and Su, T. K. (2000). Correlation between the activity of single auditory cortical neurons and sound-localization behavior in the macaque monkey. *J. Neurophysiol.* 83, 2723–2739. doi: 10.1152/jn.2000.83.5.2723

Reser, D. H., Burman, K. J., Yu, H.-H., Chaplin, T. A., Richardson, K. E., Worthy, K. H., et al. (2013). Contrasting patterns of cortical input to architectural subdivisions of the area 8 complex: a retrograde tracing study in marmoset monkeys. *Cereb. Cortex* 23, 1901–1922. doi: 10.1093/cercor/bhs177

Rohde, M., van Dam, L. C. J., and Ernst, M. O. (2016). Statistically optimal multisensory cue integration: a practical tutorial. *Multisens. Res.* 29, 279–317. doi: 10.1163/22134808-00002510

Roitman, J. D., and Shadlen, M. N. (2002). Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *J. Neurosci.* 22, 9475–9489. doi: 10.1523/jneurosci.22-21-09475.2002

Romanski, L. M. (2007). Representation and integration of auditory and visual stimuli in the primate ventral lateral prefrontal cortex. *Cereb. Cortex* 17, i61–i69. doi: 10.1093/cercor/bhm099

Romanski, L. M. (2012). Integration of faces and vocalizations in ventral prefrontal cortex: implications for the evolution of audiovisual speech. *Proc. Natl. Acad. Sci. U S A* 109, 10717–10724. doi: 10.1073/pnas.1204335109

Romanski, L. M., Bates, J. F., and Goldman-Rakic, P. S. (1999a). Auditory belt and parabelt projections to the prefrontal cortex in the rhesus monkey. *J. Comp. Neurol.* 403, 141–157. doi: 10.1002/(sici)1096-9861(19990111)403:2<141::aid-cne1>3.0.co;2-v

Romanski, L. M., Tian, B., Fritz, J., Mishkin, M., Goldman-Rakic, P. S., and Rauschecker, J. P. (1999b). Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nat. Neurosci.* 2, 1131–1136. doi: 10.1038/16056

Rudolph, K., and Pasternak, T. (1999). Transient and permanent deficits in motion perception after lesions of cortical areas MT and MST in the macaque monkey. *Cereb. Cortex* 9, 90–100. doi: 10.1093/cercor/9.1.90

Saito, H. A., Yukie, M., Tanaka, K., Hikosaka, K., Fukada, Y., and Iwai, E. (1986). Integration of direction signals of image motion in the superior temporal sulcus of the macaque monkey. *J. Neurosci.* 6, 145–157. doi: 10.1523/jneurosci.06-01-00145.1986

Salzman, C. D., and Newsome, W. T. (1994). Neural mechanisms for forming a perceptual decision. *Science* 264, 231–237. doi: 10.1126/science.8146653

Schenk, T., and Zihl, J. (1997). Visual motion perception after brain damage: I. Deficits in global motion perception. *Neuropsychologia* 35, 1289–1297. doi: 10.1016/s0028-3932(97)00004-3

Schlack, A., Sterbing-D'Angelo, S. J., Hartung, K., Hoffmann, K.-P., and Bremmer, F. (2005). Multisensory space representations in the macaque ventral intraparietal area. *J. Neurosci.* 25, 4616–4625. doi: 10.1523/JNEUROSCI.0455-05.2005

Seitz, A. R., Kim, R., and Shams, L. (2006). Sound facilitates visual learning. *Curr. Biol.* 16, 1422–1427. doi: 10.1016/j.cub.2006.05.048

Sekuler, R., Sekuler, A. B., and Lau, R. (1997). Sound alters visual motion perception. *Nature* 385:308. doi: 10.1038/385308a0

Semple, M. N., and Kitzes, L. M. (1993a). Binaural processing of sound pressure level in cat primary auditory cortex: evidence for a representation based on absolute levels rather than interaural level differences. *J. Neurophysiol.* 69, 449–461. doi: 10.1152/jn.1993.69.2.449

Semple, M. N., and Kitzes, L. M. (1993b). Focal selectivity for binaural sound pressure level in cat primary auditory cortex: two-way intensity network tuning. *J. Neurophysiol.* 69, 462–473. doi: 10.1152/jn.1993.69.2.462

Slee, S. J., and Young, E. D. (2010). Sound localization cues in the marmoset monkey. *Hear. Res.* 260, 96–108. doi: 10.1016/j.heares.2009.12.001

Smith, K. R., Okada, K., Saberi, K., and Hickok, G. (2004). Human cortical auditory motion areas are not motion selective. *Neuroreport* 15, 1523–1526. doi: 10.1097/01.wnr.0000130233.43788.4b

Smith, K. R., Saberi, K., and Hickok, G. (2007). An event-related fMRI study of auditory motion perception: no evidence for a specialized cortical system. *Brain Res.* 1150, 94–99. doi: 10.1016/j.brainres.2007.03.003

Soto-Faraco, S., Spence, C., and Kingstone, A. (2005). Assessing automaticity in the audiovisual integration of motion. *Acta Psychol.* 118, 71–92. doi: 10.1016/j.actpsy.2004.10.008

Strnad, L., Peelen, M. V., Bedny, M., and Caramazza, A. (2013). Multivoxel pattern analysis reveals auditory motion information in MT+ of both congenitally blind and sighted individuals. *PLoS One* 8:e63198. doi: 10.1371/journal.pone.0063198

Tanaka, K., and Saito, H. A. (1989). Analysis of motion of the visual field by direction, expansion/contraction, and rotation cells clustered in the dorsal part of the medial superior temporal area of the macaque monkey. *J. Neurophysiol.* 62, 626–641. doi: 10.1152/jn.1989.62.3.626

Thaler, L., Paciocco, J., Daley, M., Lesniak, G. D., Purcell, D. W., Fraser, J. A., et al. (2016). A selective impairment of perception of sound motion direction in peripheral space: a case study. *Neuropsychologia* 80, 79–89. doi: 10.1016/j.neuropsychologia.2015.11.008

Tian, B., Reser, D., Durham, A., Kustov, A., and Rauschecker, J. P. (2001). Functional specialization in rhesus monkey auditory cortex. *Science* 292, 290–293. doi: 10.1126/science.1058911

Van Essen, D. C. (2002). Surface-based atlases of cerebellar cortex in the human, macaque, and mouse. *Ann. N Y Acad. Sci.* 978, 468–479. doi: 10.1111/j.1749-6632.2002.tb07588.x

Van Essen, D. C., and Dierker, D. (2007). Surface-based and probabilistic atlases of primate cerebral cortex. *Neuron* 56, 209–225. doi: 10.1016/j.neuron.2007.10.015

von Saldern, S., and Noppeney, U. (2013). Sensory and striatal areas integrate auditory and visual signals into behavioral benefits during motion discrimination. *J. Neurosci.* 33, 8841–8849. doi: 10.1523/jneurosci.3020-12.2013

Wang, Y., Celebrini, S., Trotter, Y., and Barone, P. (2008). Visuo-auditory interactions in the primary visual cortex of the behaving monkey: electrophysiological evidence. *BMC Neurosci.* 9:79. doi: 10.1186/1471-2202-9-79

Warren, J., Zielinski, B., and Green, G. (2002). Perception of sound-source motion by the human brain. *Neuron* 34, 139–148. doi: 10.1016/s0896-6273(02)00637-2

Watanabe, J., and Iwai, E. (1991). Neuronal activity in visual, auditory and polysensory areas in the monkey temporal cortex during visual fixation task. *Brain Res. Bull.* 26, 583–592. doi: 10.1016/0361-9230(91)90099-6

Wilson, W. W., and O'Neill, W. E. (1998). Auditory motion induces directionally dependent receptive field shifts in inferior colliculus neurons. *J. Neurophysiol.* 79, 2040–2062. doi: 10.1152/jn.1998.79.4.2040

Woods, T. M., Lopez, S. E., Long, J. H., Rahman, J. E., and Recanzone, G. H. (2006). Effects of stimulus azimuth and intensity on the single-neuron activity in the auditory cortex of the alert macaque monkey. *J. Neurophysiol.* 96, 3323–3337. doi: 10.1152/jn.00392.2006

Wuerger, S. M., Hofbauer, M., and Meyer, G. F. (2003). The integration of auditory and visual motion signals at threshold. *Percept. Psychophys.* 65, 1188–1196. doi: 10.3758/bf03194844

Wuerger, S. M., Parkes, L., Lewis, P. A., Crocker-Buque, A., Rutschmann, R., and Meyer, G. F. (2012). Premotor cortex is sensitive to auditory-visual congruence for biological motion. *J. Cogn. Neurosci.* 24, 575–587. doi: 10.1162/jocn_a_00173

Yu, H.-H., and Rosa, M. G. P. (2014). Uniformity and diversity of response properties of neurons in the primary visual cortex: selectivity for orientation, direction of motion and stimulus size from center to far periphery. *Vis. Neurosci.* 31, 85–98. doi: 10.1017/s0952523813000448

Yu, H.-H., Verma, R., Yang, Y., Tibballs, H. A., Lui, L. L., Reser, D. H., et al. (2010). Spatial and temporal frequency tuning in striate cortex: functional uniformity and specializations related to receptive field eccentricity. *Eur. J. Neurosci.* 31, 1043–1062. doi: 10.1111/j.1460-9568.2010.07118.x

Zaksas, D., and Pasternak, T. (2006). Directional signals in the prefrontal cortex and in area MT during a working memory for visual motion task. *J. Neurosci.* 26, 11726–11742. doi: 10.1523/jneurosci.3420-06.2006

Zeki, S. M., Watson, J. D., Lueck, C. J., Friston, K. J., Kennard, C., and Frackowiak, R. S. (1991). A direct demonstration of functional specialization in human visual cortex. *J. Neurosci.* 11, 641–649. doi: 10.1523/jneurosci.11-03-00641.1991

Zihl, J., von Cramon, D., and Mai, N. (1983). Selective disturbance of movement vision after bilateral brain damage. *Brain* 106, 313–340. doi: 10.1093/brain/106.2.313

Understanding Sensory Information Processing Through Simultaneous Multi-area Population Recordings

*Elizabeth Zavitz [1,2] * and Nicholas S. C. Price [1,2]*

[1]*Department of Physiology, Biomedicine Discovery Institute, Monash University, Clayton, VIC, Australia,* [2]*Centre of Excellence for Integrative Brain Function, Monash University Node, Clayton, VIC, Australia*

The goal of sensory neuroscience is to understand how the brain creates its myriad of representations of the world, and uses these representations to produce perception and behavior. Circuits of neurons in spatially segregated regions of brain tissue have distinct functional specializations, and these regions are connected to form a functional processing hierarchy. Advances in technology for recording neuronal activity from multiple sites in multiple cortical areas mean that we are now able to collect data that reflects how information is transformed within and between connected members of this hierarchy. This advance is an important step in understanding the brain because, after the sensory organs have transduced a physical signal, every processing stage takes the activity of other neurons as its input, not measurements of the physical world. However, as we explore the potential of studying how populations of neurons in multiple areas respond in concert, we must also expand both the analytical tools that we use to make sense of these data and the scope of the theories that we attempt to define. In this article, we present an overview of some of the most promising analytical approaches for making inferences from population recordings in multiple brain areas, such as dimensionality reduction and measuring changes in correlated variability, and examine how they may be used to address longstanding questions in sensory neuroscience.

Keywords: neuronal populations, hierarchical processing, neural computation, sensory coding, inter-area communication

INTRODUCTION

The cortex contains a multitude of representations of sensory information that are anatomically segregated by sensory modality (e.g., somatosensory vs. auditory), and by specialty within a modality (e.g., visual motion vs. visual form). Following recent advances in technology, large-scale recordings of neuronal population activity now extend across the boundaries of cortical areas. This presents an opportunity to understand the nature of inter-area neural processing. Many inter-neuronal and inter-area phenomena exist on timescales of milliseconds. In order to characterize this short-timescale activity requires electrophysiological approaches, which allow action potentials and local field potentials (LFPs) to be recorded. Although the largest simultaneous recordings of the functional activity of neuronal ensembles are now conducted with cellular-resolution imaging, and while cell-type specific genetic promoters promise recordings from neurons with known classes (Luo et al., 2008), in this article we will focus on experiments involving extracellular electrophysiological measurements, because these afford the temporal resolution required to address the analytical questions we pose. We mainly consider cortico-cortical processing in non-human primates, but these advances are complemented by substantial work in other species,

and involving sub-cortical areas, which will be necessary to bridge the gap between understanding circuit architecture and large-scale network dynamics. Cortico-cortical processing is a good first frontier in multi-area population analysis as cortical architecture is well-characterized and similar between brain areas. Further, we mainly consider questions pertinent to data sets with population recordings from multiple brain areas simultaneously, but draw inspiration from analytical methods applied to either population recordings from one brain area, or recordings of two units in different areas.

WHY AND HOW SHOULD WE MAKE SIMULTANEOUS MULTI-AREA POPULATION RECORDINGS?

The transition from recording from a single site at one time to recording population activity was a meaningful one for systems electrophysiology (Brown et al., 2004; Yuste, 2015). Recording from populations allows us to "embrace single-neuron heterogeneity" (Cunningham and Yu, 2014), and reveals structure in the signals across multiple neurons that we would not be able to recover any other way, such as their correlated variability (Zavitz et al., 2016; Bondy et al., 2018), and how population representations change within a subspace over time or depending on context (Churchland et al., 2012). Recording simultaneously from two or more neurons has advanced theories relating to how different types of "noise," or inter-trial variability, affect stimulus discrimination (Zohary et al., 1994; Shadlen and Newsome, 1998; Cohen and Kohn, 2011; Kohn et al., 2016), and how decisions are generated based on the accumulation of evidence (Yates et al., 2017).
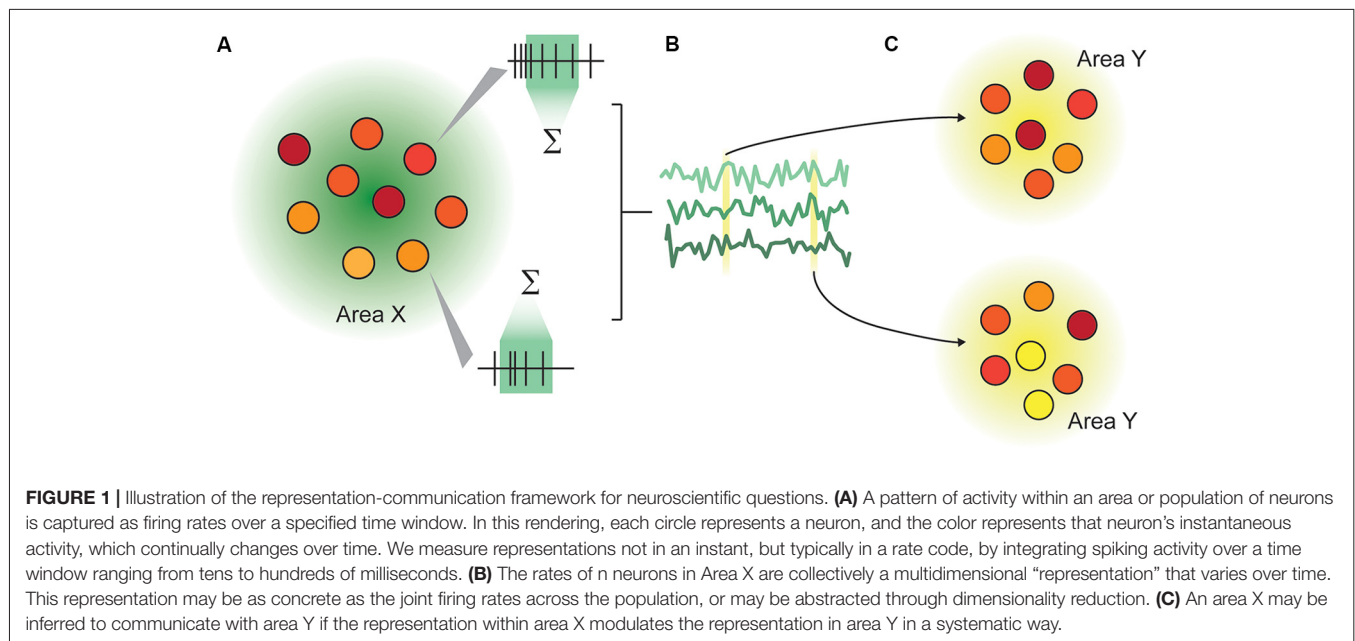
Recording from multiple areas can reveal temporal correlations between the two areas (Wong et al., 2016), giving insight into inter-area connectivity. Beyond this, by making simultaneous multi-area population recordings, we are able to make inferences about how population representations in one area influence the representations in another on a trial-by-trial basis (Zandvakili and Kohn, 2015), and how inter-area communication changes depending on external factors such as attention (Ruff and Cohen, 2016). Multi-area population recordings are thus able to address two classes of questions: how representations are changed between cortical areas, and how communication is facilitated (**Figure 1**). Here, representations are defined as the structure of neuronal activity in an ensemble, and communication as a recoding process (Pitkow and Angelaki, 2017), in which the representation of information within the recipient area is measurably changed. A similar architecture is outlined in Fries (2015).

Most sensory neuroscience is predicated on developing an understanding of how a physical stimulus produces an observed neuronal response. However, beyond the level of our sensory receptors, neurons do not directly respond to sensory stimuli. Rather, they change their membrane potential and generate action potentials in response to precise patterns of inputs, received from a population of synaptically-connected neurons. By recording from connected brain areas, we can use the recordings from the source area to gain a better understanding of the true inputs to the recipient brain area, and how they are transformed in the downstream area.

PROMISING ANALYTICAL APPROACHES

There are three major classes of analyses that have allowed researchers to draw novel conclusions about information processing between simultaneously recorded areas: lower-dimensional representations; pairwise correlated variability ("noise" correlations or "correlation structure"); and measures

FIGURE 1 | Illustration of the representation-communication framework for neuroscientific questions. **(A)** A pattern of activity within an area or population of neurons is captured as firing rates over a specified time window. In this rendering, each circle represents a neuron, and the color represents that neuron's instantaneous activity, which continually changes over time. We measure representations not in an instant, but typically in a rate code, by integrating spiking activity over a time window ranging from tens to hundreds of milliseconds. **(B)** The rates of n neurons in Area X are collectively a multidimensional "representation" that varies over time. This representation may be as concrete as the joint firing rates across the population, or may be abstracted through dimensionality reduction. **(C)** An area X may be inferred to communicate with area Y if the representation within area X modulates the representation in area Y in a systematic way.
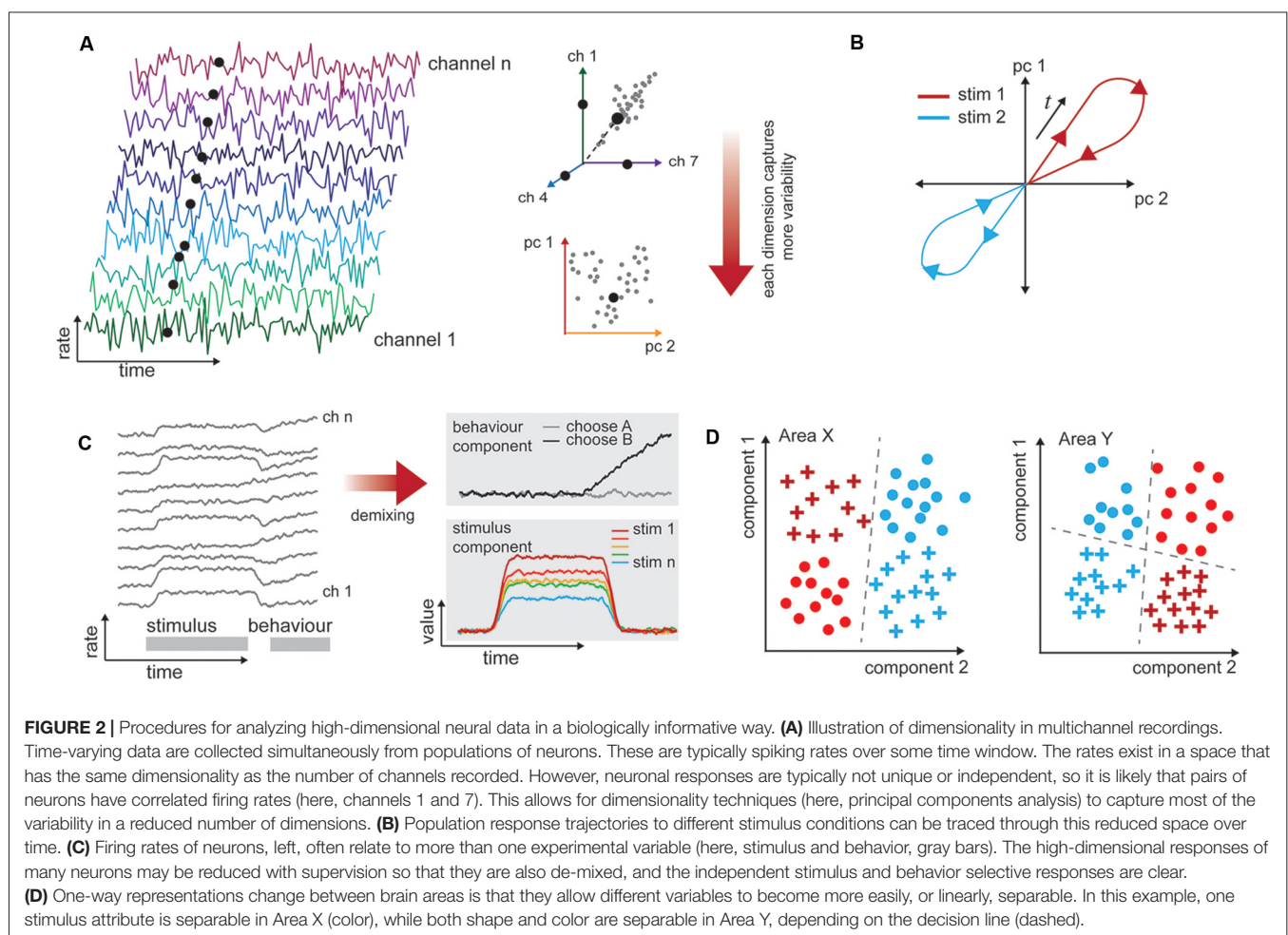
of spike-timing precision. The most valuable observations we derive from these analyses are often not their immediate outputs, but instead how these outputs change depending on other contextual variables such as the stimulus, behavior, or cognitive state.

Lower-Dimensional Representations

Across a population of neurons, there is both diversity and redundancy in neuronal responses, and it can be difficult to gain any understanding of how sensory information is represented when the number of dimensions describing the data equals the observed number of neurons (**Figure 2A**). Dimensionality reduction techniques such as principal components analysis allow covariation between neurons to be collapsed (**Figure 2B**), and the resulting visualization can show how population representations shift as a function of time and stimulus properties (**Figure 2C**). By translating data into a reduced format, we can form intuitions and hypotheses about what would otherwise be an intractably large data set that may bear little relationship to stimulus variables at first examination (Cunningham and Yu, 2014). In this "state space" the aggregate population activity at any point in time may be represented by a single point (**Figure 2B**). This style of

representation permits comparison across stimulus or behavioral characteristics independently of the often heterogeneous and complex selectivity of the neurons (as in Churchland et al., 2012; Mante et al., 2013). Dimensionality reduction can be achieved in a number of ways (principal components analysis, factor analysis, Gaussian process factor analysis, among others), with different methodological advantages but similar outcome: a reduced space in which to consider the variability of neuronal responses. Traditionally, the focus is on how this variability relates to the stimulus or behavior. With multi-area recordings, it is also appropriate to consider how the variability of neuronal responses in one area relates to the responses of a connected population.

In a typical experiment in which multiple factors can vary (e.g., stimulus value, animal behavioral state, motor outcome), the variability in neuronal responses across trials of the same type is the most interesting to the experimenter. Unsupervised approaches will operate on the data blind to these experimental manipulations or outcomes, and the components they extract may not isolate the impact of experimental variables of interest (Kobak et al., 2016). To address this shortcoming, a layer of supervision can be added to isolate experimental variables, e.g., hierarchical decomposition (Repucci et al., 2001;

FIGURE 2 | Procedures for analyzing high-dimensional neural data in a biologically informative way. **(A)** Illustration of dimensionality in multichannel recordings. Time-varying data are collected simultaneously from populations of neurons. These are typically spiking rates over some time window. The rates exist in a space that has the same dimensionality as the number of channels recorded. However, neuronal responses are typically not unique or independent, so it is likely that pairs of neurons have correlated firing rates (here, channels 1 and 7). This allows for dimensionality techniques (here, principal components analysis) to capture most of the variability in a reduced number of dimensions. **(B)** Population response trajectories to different stimulus conditions can be traced through this reduced space over time. **(C)** Firing rates of neurons, left, often relate to more than one experimental variable (here, stimulus and behavior, gray bars). The high-dimensional responses of many neurons may be reduced with supervision so that they are also de-mixed, and the independent stimulus and behavior selective responses are clear. **(D)** One-way representations change between brain areas is that they allow different variables to become more easily, or linearly, separable. In this example, one stimulus attribute is separable in Area X (color), while both shape and color are separable in Area Y, depending on the decision line (dashed).

Maddess et al., 2006), demixed PCA (Kobak et al., 2016), and tensor component analysis (Williams et al., 2017). This means that the recovered components are those that best explain individual and paired factors of interest (Brendel et al., 2011; Kobak et al., 2016). We illustrate a simplified account of mixed "stimulus" and "behavior" signals in a population, and how these components may appear once demixed in **Figure 2D**. Although poorly explored thus far, we anticipate that this approach will be particularly valuable for analyzing multi-area data sets, because it will enable quantification of how the representations change together on a trial-by-trial basis.

Dimensionality reduction works by collapsing across shared variability that arises from variations in both the "signal" (i.e., tuning similarities) and the "noise" (i.e., trial-by-trial variations in responses to the same signal). To learn more about the nature of population representations and inter-area communication, we can examine the noise correlations in isolation.

Noise Correlations

The spiking activity of neurons varies from trial to trial, even under identical stimulation conditions. In pairs of simultaneously recorded neurons, this variability tends to be shared: if one neuron fires at an above-average rate, others are likely to as well (Zohary et al., 1994). Because this shared variability is not related to the stimulus or signal, it is termed "noise" or "spike-count" correlations, and is quantified by the Pearson's correlation coefficient between the spike counts of the two cells across repetitions of the same stimulus (Cohen and Kohn, 2011). The strength of the measured correlation depends on a number of factors, including the two neurons' mean firing rate (de la Rocha et al., 2007), separation in cortical tissue (Smith and Kohn, 2008; Solomon et al., 2015; Rosenbaum et al., 2017), and similarity in tuning properties (Kohn and Smith, 2005).

The pattern of spike-count correlations we are able to observe can reflect global modulations in activity that affect the whole population (Goris et al., 2014) or synaptic architecture, which can describe either structural architecture like connectivity patterns (Hu et al., 2012) or functional architecture like moment-to-moment connectivity (Haider and McCormick, 2009). Functional architecture, and spike-count correlations, are changed by recruiting (Snyder et al., 2014) or adapting (Zavitz et al., 2016) different subpopulations of neurons. The magnitude and structure of pairwise correlated variability across populations of neurons relates to behavior (Gutnisky et al., 2017; Ni et al., 2018), how well stimulus parameters are represented (Moreno-Bote et al., 2014; Kohn et al., 2016; Zylberberg et al., 2016; Zavitz et al., 2017), and reflects the task the animal is performing (Bondy et al., 2018).

To measure spike-count correlations, spikes are typically counted in bins with sizes ranging from tens of milliseconds to one or two seconds. However, information is also present in the precise timing of spikes from a neuron, either relative to the LFP or the timing of spikes from other neurons. While longer bins increase the overall spike count and the reliability of

the measure, the behavioral relevance of these timescales is not clear.

Spike-Timing Precision

The precise timing relationships in the activity of groups of neurons, measured as synchrony or coherence, can inform us about coordinated spiking activity and communication (Jia et al., 2013; Zandvakili and Kohn, 2015). Synchronized firing across a diverse group of neurons may be an important way to encode complex stimuli (Singer et al., 1997), and pairs of neurons can coordinate firing at timescales as short as 1 ms (Palm et al., 1988). There is evidence that different information is encoded in spikes aligned with different phases of specific frequencies in the LFP (Womelsdorf et al., 2012; Wong et al., 2016) and neural activity with precise delays between populations of neurons and across cortical layers may even be critical to the process of information transmission (Bastos et al., 2015).

Spiking synchrony may be measured with a cross-correlogram—correlations in instantaneous spiking between neurons at a range of time delays. While spiking activity is best understood as a point-process in the time domain, the LFP is a continuous process in the time-frequency domain, characterized in terms of how the power and phase across different frequency bands change over time. A common way of relating these discrete and continuous processes is coherence, a frequency-dependent measure of signal correlation, that may be examined between spikes and the LFP recorded on the same or different electrodes (Jarvis and Mitra, 2001). These measures have been used to understand how pairs of neurons communicate within (Dean et al., 2012; Hagan et al., 2012) and between (Jia et al., 2013; Wong et al., 2016) cortical areas. Although their use has not yet been expanded to large-scale recordings, given that spikes are commonly described as the outputs of a neuron and the LFP represents the net synaptic input to the region near the electrode, these approaches correlating spiking and the LFP are some of the most direct for examining how communication occurs across area boundaries. There are not any widely adopted population measures of timing precision, and this presents a fruitful area for future development. The process of identifying assemblies of neurons that fire in concert (Singer et al., 1997) could be expanded to include more detailed temporal characterization.

VIABLE AVENUES OF INQUIRY

How Do Brain Areas Communicate With One Another?

Information is flexibly and efficiently routed throughout the brain. Here, we define communication as signal propagation that produces a change in the representation by a recipient area. Part of the challenge for achieving inter-area communication is related to signal transmission: a signal must be able propagate reliably throughout the system without excessive attenuation or amplification (Shadlen and Newsome, 1998; Joglekar et al., 2018; van Vugt et al., 2018). This relies on inter-area anatomical connections as well as the network structure within an

area (Joglekar et al., 2018). However, there is substantial evidence that successful inter-area communication also requires physiological coordination on millisecond time-scales (Fries, 2005, 2015).

Inter-area information transmission has been assessed using coherence measures across the V1-V2 boundary (Jia et al., 2013), and by the likelihood of spikes in a recipient area given the state of a source area (Zandvakili and Kohn, 2015). The quality of signal transmission has been measured by the number of spikes elicited in the recipient area following of electrostimulation of the source area (Ruff and Cohen, 2016). These approaches demonstrate an effect of state on a recipient area, or propagation, but they do not demonstrate that communication has occurred. This could be achieved with an additional analysis demonstrating improvement in coding in the recipient brain area. This may be done directly by assessing perception in an awake behaving animal or decoding the spiking activity in the anesthetized preparation; or indirectly by measuring representations or spike-count correlations. These early studies had a small number of electrodes in the recipient area, so such analyses would have been limited, but will be increasingly possible as recording capabilities improve. Changes in noise correlations between areas can also be interpreted as changes in the communication efficacy between areas. If correlations between areas increase, they share more trial-to-trial variability, which means signal transmission is enhanced, but it is unknown whether this also enhances the representation in the recipient brain area.

Within a single brain area, inferences may be made about the relationship between cortical state and coding efficacy by conditioning the data, or sorting population activity into states based on a variable of interest (e.g., up and down states based on firing rate; Arandia-Romero et al., 2016; Gutnisky et al., 2017), or behavioral outcome or strategy (Gilad et al., 2018). Recent work adapts this approach to two connected populations of neurons by estimating how the state of one area impacts coding in a recipient area, demonstrating how we might test the efficacy of neural communication more directly (Palmigiano et al., 2017). In simulations, they measured the relative phase of gamma bursts in two areas, and condition based on which area is leading. This enabled them to show that spiking activity in the leading area predicts spiking activity in the following area, suggesting that gamma bursts produce states that are conducive to spike transmission. However, the results of conditioning data should be interpreted with caution, as the variable chosen for conditioning will have multiple covariates.

How Are Representations Transformed Between Areas?

Understanding population responses in terms of a low-dimensional representation has provided traction especially in our understanding of how neurons with complex selectivity represent stimuli and guide behavior. In the context of multi-area recordings, this approach stands to help us understand how representations of the same factors shift from one area to another, and how shifts in the trial-by-trial activity in an upstream area produce better or worse representations in a downstream area. It also provides a way to look at how different areas reshape the same information in order to "untangle" it, or increase the linear separability of a biologically relevant variable (**Figure 2D**; DiCarlo and Cox, 2007; DiCarlo et al., 2012; Pagan et al., 2013). In future work, dimensionality reduction may be combined with data conditioning in order to determine how the representation in a recipient area depends on the state of a simultaneously recorded source area.

This problem extends to reasoning about how different areas contribute to different aspects of a complex task. Yates et al. (2017) combined measurements of behavior and the spike-count correlations within and between areas MT and LIP, with models of the two areas. They were able to dissect a perceptual decision-making task into several components that are partially shared between MT and LIP, but did not find any evidence of single-trial coupling between these two areas, which is inconsistent with theories that LIP integrates the information in MT. Simultaneous population recordings in multiple areas alone permit this kind of trial-by-trial assessment of how information is transferred and transformed, and will be useful for separating hierarchical computations from computations that are apparent at many stages of the hierarchy.

How Do Global Factors Modulate Inter-Area Cortico-Cortical Communication?

Variability in the responses of neurons, as measured with spike-count correlations, can be partly explained by modulating factors such as anesthetic state, attention, and arousal (Goris et al., 2014; Rabinowitz et al., 2015). It is unclear how these "global" factors interact with local factors (such as adaptation or stimulus context), and what the scale of the modulations induced by these global factors truly is. By recording population activity in multiple areas, we will be able to determine the scope of local and global factors, for example, to determine how far local network changes propagate through the cortical hierarchy. Sub-cortical systems play a significant role in modulating cortical processing (Sherman, 2016). Expanding simultaneous multi-area cortical recordings to include related subcortical systems, potentially in small brains with large, multi-contact probes (Jun et al., 2017), may be profoundly informative for learning why cortical states tend to shift, both "spontaneously" and in a task-dependent way (Ruff and Cohen, 2018).

CONCLUSION

We are able to measure larger populations than ever, but characterizing many predicted theoretical effects requires recording from exceedingly large-scale populations (hundreds or thousands of neurons). While most electrophysiology is currently constrained to monitoring hundreds of neurons, imaging approaches are able to monitor thousands but have poor temporal resolution. Improved temporal resolution of imaging and higher-yield electrophysiology experiments will move the field forward substantially.

Population size aside, dimensionality reduction requires repeating each trial a large number of times (and indeed, the number of necessary repetitions increases with the number of cells simultaneously recorded). The recording stability required for these measurements can be difficult to obtain in an anesthetized preparation and the timescale is potentially impossible in awake animals until recordings can be reconciled with carefully quantified natural behaviors. In single-area recordings, the limits of the anesthetized preparation are reasonably well-understood, but it is not yet clear if inter-area dynamics are as consistent as basic sensory representations between the anesthetized and awake states. Modest increases in population size, along with the technological advances that permit us to characterize each cell more completely (e.g., laminar profile, genetic markers, morphology, receptive field substructures, connectivity) will let us make stronger inferences about the varied roles different cells play in shaping population activity, and thus perception, cognition, and behavior.

AUTHOR CONTRIBUTIONS

EZ and NP wrote and edited the article.

FUNDING

REFERENCES

Arandia-Romero, I., Tanabe, S., Drugowitsch, J., Kohn, A., and Moreno-Bote, R. (2016). Multiplicative and additive modulation of neuronal tuning with population activity affects encoded information. *Neuron* 89, 1305–1316. doi: 10.1016/j.neuron.2016.01.044

Bastos, A. M., Vezoli, J., and Fries, P. (2015). Communication through coherence with inter-areal delays. *Curr. Opin. Neurobiol.* 31, 173–180. doi: 10.1016/j.conb.2014.11.001

Bondy, A. G., Haefner, R. M., and Cumming, B. G. (2018). Feedback determines the structure of correlated variability in primary visual cortex. *Nat. Neurosci.* 21, 598–606. doi: 10.1038/s41593-018-0089-1

Brendel, W., Romo, R., and Machens, C. K. (2011). "Demixed principal component analysis," in *Advances in Neural Information Processing Systems 24*, eds J. Shawe-Taylor, R. S. Zemel, P. L. Bartlett, F. Pereira and K. Q. Weinberger (Granada: Curran Associates, Inc.), 2654–2662.

Brown, E. N., Kass, R. E., and Mitra, P. P. (2004). Multiple neural spike train data analysis: state-of-the-art and future challenges. *Nat. Neurosci.* 7, 456–461. doi: 10.1038/nn1228

Churchland, M. M., Cunningham, J. P., Kaufman, M. T., Foster, J. D., Nuyujukian, P., Ryu, S. I., et al. (2012). Neural population dynamics during reaching. *Nature* 487, 51–56. doi: 10.1038/nature11129

Cohen, M. R., and Kohn, A. (2011). Measuring and interpreting neuronal correlations. *Nat. Neurosci.* 14, 811–819. doi: 10.1038/nn.2842

Cunningham, J. P., and Yu, B. M. (2014). Dimensionality reduction for large-scale neural recordings. *Nat. Neurosci.* 17, 1500–1509. doi: 10.1038/nn.3776

de la Rocha, J., Doiron, B., Shea-Brown, E., Josić, K., and Reyes, A. (2007). Correlation between neural spike trains increases with firing rate. *Nature* 448, 802–806. doi: 10.1038/nature06028

Dean, H. L., Hagan, M. A., and Pesaran, B. (2012). Only coherent spiking in posterior parietal cortex coordinates looking and reaching. *Neuron* 73, 829–841. doi: 10.1016/j.neuron.2011.12.035

DiCarlo, J. J., and Cox, D. D. (2007). Untangling invariant object recognition. *Trends Cogn. Sci.* 11, 333–341. doi: 10.1016/j.tics.2007.06.010

DiCarlo, J. J., Zoccolan, D., and Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron* 73, 415–434. doi: 10.1016/j.neuron.2012.01.010

Fries, P. (2005). A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends Cogn. Sci.* 9, 474–480. doi: 10.1016/j.tics.2005.08.011

Fries, P. (2015). Rhythms for cognition: communication through coherence. *Neuron* 88, 220–235. doi: 10.1016/j.neuron.2015.09.034

Gilad, A., Gallero-Salas, Y., Groos, D., and Helmchen, F. (2018). Behavioral strategy determines frontal or posterior location of short-term memory in neocortex. *Neuron* 99, 814.e7–828.e7. doi: 10.1016/j.neuron.2018.07.029

Goris, R. L. T., Movshon, J. A., and Simoncelli, E. P. (2014). Partitioning neuronal variability. *Nat. Neurosci.* 17, 858–865. doi: 10.1038/nn.3711

Gutnisky, D. A., Beaman, C., Lew, S. E., and Dragoi, V. (2017). Cortical response states for enhanced sensory discrimination. *Elife* 6:e29226. doi: 10.7554/eLife.29226

Hagan, M. A., Dean, H. L., and Pesaran, B. (2012). Spike-field activity in parietal area LIP during coordinated reach and saccade movements. *J. Neurophysiol.* 107, 1275–1290. doi: 10.1152/jn.00867.2011

Haider, B., and McCormick, D. A. (2009). Rapid neocortical dynamics: cellular and network mechanisms. *Neuron* 62, 171–189. doi: 10.1016/j.neuron.2009.04.008

Hu, Y., Trousdale, J., Josić, K., and Shea-Brown, E. (2012). Motif statistics and spike correlations in neuronal networks. *BMC Neurosci.* 13:P43. doi: 10.1186/1471-2202-13-s1-p43

Jarvis, M. R., and Mitra, P. P. (2001). Sampling properties of the spectrum and coherency of sequences of action potentials. *Neural Comput.* 13, 717–749. doi: 10.1162/089976601300014312

Jia, X., Tanabe, S., and Kohn, A. (2013). Gamma and the coordination of spiking activity in early visual cortex. *Neuron* 77, 762–774. doi: 10.1016/j.neuron.2012.12.036

Joglekar, M. R., Mejias, J. F., Yang, G. R., and Wang, X.-J. (2018). Inter-areal balanced amplification enhances signal propagation in a large-scale circuit model of the primate cortex. *Neuron* 98, 222.e8–234.e8. doi: 10.1016/j.neuron.2018.02.031

Jun, J. J., Steinmetz, N. A., Siegle, J. H., Denman, D. J., Bauza, M., Barbarits, B., et al. (2017). Fully integrated silicon probes for high-density recording of neural activity. *Nature* 551, 232–236. doi: 10.1038/nature24636

Kobak, D., Brendel, W., Constantinidis, C., Feierstein, C. E., Kepecs, A., Mainen, Z. F., et al. (2016). Demixed principal component analysis of neural population data. *Elife* 5:e10989. doi: 10.7554/eLife.10989

Kohn, A., Coen-Cagli, R., Kanitscheider, I., and Pouget, A. (2016). Correlations and neuronal population information. *Annu. Rev. Neurosci.* 39, 237–256. doi: 10.1146/annurev-neuro-070815-013851

Kohn, A., and Smith, M. A. (2005). Stimulus dependence of neuronal correlation in primary visual cortex of the macaque. *J. Neurosci.* 25, 3661–3673. doi: 10.1523/JNEUROSCI.5106-04.2005

Luo, L., Callaway, E. M., and Svoboda, K. (2008). Genetic dissection of neural circuits. *Neuron* 57, 634–660. doi: 10.1016/j.neuron.2008.01.002

Maddess, T., James, A. C., Ruseckaite, R., and Bowman, E. A. (2006). Hierarchical decomposition of dichoptic multifocal visual evoked potentials. *Vis. Neurosci.* 23, 703–712. doi: 10.1017/s0952523806230013

Mante, V., Sussillo, D., Shenoy, K. V., and Newsome, W. T. (2013). Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature* 503, 78–84. doi: 10.1038/nature12742

Moreno-Bote, R., Beck, J., Kanitscheider, I., Pitkow, X., Latham, P., and Pouget, A. (2014). Information-limiting correlations. *Nat. Neurosci.* 17, 1410–1417. doi: 10.1038/nn.3807

Ni, A. M., Ruff, D. A., Alberts, J. J., Symmonds, J., and Cohen, M. R. (2018). Learning and attention reveal a general relationship between population activity and behavior. *Science* 359, 463–465. doi: 10.1126/science.aao0284

Pagan, M., Urban, L. S., Wohl, M. P., and Rust, N. C. (2013). Signals in inferotemporal and perirhinal cortex suggest an untangling of visual target information. *Nat. Neurosci.* 16, 1132–1139. doi: 10.1038/nn.3433

Palm, G., Aertsen, A. M., and Gerstein, G. L. (1988). On the significance of correlations among neuronal spike trains. *Biol. Cybern.* 59, 1–11. doi: 10.1007/bf00336885

Palmigiano, A., Geisel, T., Wolf, F., and Battaglia, D. (2017). Flexible information routing by transient synchrony. *Nat. Neurosci.* 20, 1014–1022. doi: 10.1038/nn.4569

Pitkow, X., and Angelaki, D. E. (2017). Inference in the brain: statistics flowing in redundant population codes. *Neuron* 94, 943–953. doi: 10.1016/j.neuron.2017.05.028

Rabinowitz, N. C., Goris, R. L., Cohen, M., and Simoncelli, E. P. (2015). Attention stabilizes the shared gain of V4 populations. *Elife* 4:e08998. doi: 10.7554/eLife.08998

Repucci, M. A., Schiff, N. D., and Victor, J. D. (2001). General strategy for hierarchical decomposition of multivariate time series: implications for temporal lobe seizures. *Ann. Biomed. Eng.* 29, 1135–1149. doi: 10.1114/1.1424914

Rosenbaum, R., Smith, M. A., Kohn, A., Rubin, J. E., and Doiron, B. (2017). The spatial structure of correlated neuronal variability. *Nat. Neurosci.* 20, 107–114. doi: 10.1038/nn.4433

Ruff, D. A., and Cohen, M. R. (2016). Attention increases spike count correlations between visual cortical areas. *J. Neurosci.* 36, 7523–7534. doi: 10.1523/JNEUROSCI.0610-16.2016

Ruff, D. A., and Cohen, M. R. (2018). Simultaneous multi-area recordings suggest a novel hypothesis about how attention improves performance. *bioRxiv:372888 [Preprint].* doi: 10.1101/372888

Shadlen, M. N., and Newsome, W. T. (1998). The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *J. Neurosci.* 18, 3870–3896. doi: 10.1523/JNEUROSCI.18-10-03870.1998

Sherman, S. M. (2016). Thalamus plays a central role in ongoing cortical functioning. *Nat. Neurosci.* 19, 533–541. doi: 10.1038/nn.4269

Singer, W., Engel, A. K., Kreiter, A. K., Munk, M. H., Neuenschwander, S., and Roelfsema, P. R. (1997). Neuronal assemblies: necessity, signature and detectability. *Trends Cogn. Sci.* 1, 252–261. doi: 10.1016/S1364-6613(97)01079-6

Smith, M. A., and Kohn, A. (2008). Spatial and temporal scales of neuronal correlation in primary visual cortex. *J. Neurosci.* 28, 12591–12603. doi: 10.1523/JNEUROSCI.2929-08.2008

Snyder, A. C., Morais, M. J., Kohn, A., and Smith, M. A. (2014). Correlations in V1 are reduced by stimulation outside the receptive field. *J. Neurosci.* 34, 11222–11227. doi: 10.1523/JNEUROSCI.0762-14.2014

Solomon, S. S., Chen, S. C., Morley, J. W., and Solomon, S. G. (2015). Local and global correlations between neurons in the middle temporal area of primate visual cortex. *Cereb. Cortex* 25, 3182–3196. doi: 10.1093/cercor/bhu111

van Vugt, B., Dagnino, B., Vartak, D., Safaai, H., Panzeri, S., Dehaene, S., et al. (2018). The threshold for conscious report: signal loss and response

bias in visual and frontal cortex. *Science* 360, 537–542. doi: 10.1126/science.aar7186

Williams, A. H., Kim, T. H., Wang, F., Vyas, S., Ryu, S. I., Shenoy, K. V., et al. (2017). Unsupervised discovery of demixed, low-dimensional neural dynamics across multiple timescales through tensor components analysis. *bioRxiv:211128 [Preprint].* doi: 10.1101/211128

Womelsdorf, T., Lima, B., Vinck, M., Oostenveld, R., Singer, W., Neuenschwander, S., et al. (2012). Orientation selectivity and noise correlation in awake monkey area V1 are modulated by the gamma cycle. *Proc. Natl. Acad. Sci. U S A* 109, 4302–4307. doi: 10.1073/pnas.1114223109

Wong, Y. T., Fabiszak, M. M., Novikov, Y., Daw, N. D., and Pesaran, B. (2016). Coherent neuronal ensembles are rapidly recruited when making a look-reach decision. *Nat. Neurosci.* 19, 327–334. doi: 10.1038/nn.4210

Yates, J. L., Park, I. M., Katz, L. N., Pillow, J. W., and Huk, A. C. (2017). Functional dissection of signal and noise in MT and LIP during decision-making. *Nat. Neurosci.* 20, 1285–1292. doi: 10.1038/nn.4611

Yuste, R. (2015). From the neuron doctrine to neural networks. *Nat. Rev. Neurosci.* 16, 487–497. doi: 10.1038/nrn3962

Zandvakili, A., and Kohn, A. (2015). Coordinated neuronal activity enhances corticocortical communication. *Neuron* 87, 827–839. doi: 10.1016/j.neuron.2015.07.026

Zavitz, E., Yu, H.-H., Rosa, M. G. P., and Price, N. S. C. (2017). Correlated variability in the neurons with the strongest tuning improves direction coding. *Cereb. Cortex* doi: 10.1093/cercor/bhx344 [Epub ahead of print].

Zavitz, E., Yu, H.-H., Rowe, E. G., Rosa, M. G. P., and Price, N. S. C. (2016). Rapid adaptation induces persistent biases in population codes for visual motion. *J. Neurosci.* 36, 4579–4590. doi: 10.1523/JNEUROSCI.4563-15.2016

Zohary, E., Shadlen, M. N., and Newsome, W. T. (1994). Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature* 370, 140–143. doi: 10.1038/370140a0

Zylberberg, J., Cafaro, J., Turner, M. H., Shea-Brown, E., and Rieke, F. (2016). Direction-selective circuits shape noise to ensure a precise population code. *Neuron* 89, 369–383. doi: 10.1016/j.neuron.2015.11.019

Probabilistic Encoding Models for Multivariate Neural Data

Marcus A. Triplett and Geoffrey J. Goodhill *

Queensland Brain Institute and School of Mathematics and Physics, The University of Queensland, Brisbane, QLD, Australia

A key problem in systems neuroscience is to characterize how populations of neurons encode information in their patterns of activity. An understanding of the encoding process is essential both for gaining insight into the origins of perception and for the development of brain-computer interfaces. However, this characterization is complicated by the highly variable nature of neural responses, and thus usually requires probabilistic methods for analysis. Drawing on techniques from statistical modeling and machine learning, we review recent methods for extracting important variables that quantitatively describe how sensory information is encoded in neural activity. In particular, we discuss methods for estimating receptive fields, modeling neural population dynamics, and inferring low dimensional latent structure from a population of neurons, in the context of both electrophysiology and calcium imaging data.

Keywords: neural coding, calcium imaging, population code, brain-computer interfaces, generalized linear model, Gaussian process, factor analysis

1. INTRODUCTION

An animal's perceptual capabilities critically depend on the ability of its brain to form appropriate representations of sensory stimuli. However, the neural activity induced by a specific stimulus is highly variable, suggesting that neural encoding is a fundamentally probabilistic process. Characterizing the neural code thus requires statistical methods for relating stimuli to distributions of evoked patterns of activity. Modern techniques for recording such neural activity include multi-electrode arrays, which provide access to the behavior of populations of neurons at millisecond resolution, and optical imaging with genetically encoded calcium (Chen et al., 2013) and voltage indicators (Abdelfattah et al., 2018), which allow thousands of neurons to be recorded simultaneously (Ahrens et al., 2013; Chen et al., 2018). However, while improvements in multi-neuron recording allow us to probe neural circuits in great detail, they are accompanied by a need for computational techniques that scale to entire neural populations.

A statistical model for neural coding describes how a stimulus is mathematically related to a pattern of neural activity. By fitting the model one can extract important variables that quantitatively describe the encoding procedure taking place. For instance, such models enable the estimation of receptive fields and/or interneuronal coupling strengths. In contrast to other methods for inferring these variables, an approach based on statistical models situates the task of estimating salient parameters in a coherent mathematical framework, often with proof of asymptotic optimality or computational efficiency. By making explicit assumptions about how the data was generated, statistically principled approaches are often capable of identifying patterns in neural data which are challenging to find with simpler methods.

Linear and generalized linear models are among the most straightforward classes of statistical models for spike trains and assume that a neuron's activity is a noisy linear combination of the stimulus features. These models are highly effective at explaining the structure of sensory receptive fields and are computationally tractable, but do not explicitly model the temporal structure of the recorded signal and have difficulty accounting for correlations between neurons in short time windows. An important aspect of these correlations is their tendency to be modular, with distinct groups of neurons showing cofluctuating activity. Latent factor models attempt to uncover the low dimensional structure that gives rise to this correlated variability, and recent efforts have focused on extracting low dimensional structure that evolves smoothly through time using a latent linear dynamical system or Gaussian process (Cunningham and Byron, 2014).

A further challenge is presented by calcium imaging, which provides only indirect access to neural activity through recorded fluorescence levels that reflect the concentration of calcium within a neuron. Often this data can be more difficult to interpret than electrophysiological recordings as there are a number of biophysical stages between stimulus presentation and fluorescence imaging where noise can enter and information can be lost. Using a generative model for calcium imaging data, however, one can explicitly account for the process through which action potentials are transformed into fluorescence levels. Fitting the generative model amounts to deconvolving the fluorescence signal to estimate the underlying spike train timeseries, and conventional encoding models can then be applied to deconvolved data. However, the ability to obtain spike counts from fluorescence data is highly constrained by experimental conditions, which motivates the development of encoding models specific to calcium imaging that do not necessarily involve spike train deconvolution.

While previous reviews have focused on estimating stimulus-response functions (Paninski et al., 2007; Pillow, 2007; Meyer et al., 2017), neural decoding (Paninski et al., 2007; Quiroga and Panzeri, 2009), and conceptual overviews of models and data analysis techniques (Cunningham and Byron, 2014; Paninski and Cunningham, 2018), this review instead discusses a range of recent exemplary models and their successful application to experimental data. Our goal is to provide sufficient mathematical detail to appreciate the respective strengths and weaknesses of each model, while leaving formal treatment of their associated fitting algorithms to their original sources.

2. LINEAR AND GENERALIZED LINEAR MODELS

We first briefly review now-standard material on models for single-neuron spike trains, primarily to develop the theory, terminology, and notation necessary for more recent work focused on multivariate models.

2.1. The Linear-Gaussian Model

Among the simplest probabilistic models for a neuron's response r to a stimulus vector $\mathbf{s}$ is the linear-Gaussian model (**Figure 1A**), which assumes that a neuron linearly filters the features of $\mathbf{s}$ as

$$r = \mathbf{w}^\top \mathbf{s} + \epsilon, \qquad \epsilon \sim \mathcal{N}(0, \sigma^2) \tag{1}$$

where the vector $\mathbf{w}$ is the stimulus filter, ϵ is an additive noise variable, and $\mathcal{N}(0, \sigma^2)$ is a Gaussian distribution with mean 0 and variance σ^2 (see **Table 1** for a table of notation). In the case of visual processing the stimulus $\mathbf{s}$ is a vector of pixel intensities for each point in the visual field, the stimulus filter $\mathbf{w}$ corresponds to the classical visual receptive field, and the response r is either the spike count or firing rate within some time window following the stimulus. Assuming stimuli $\mathbf{s}_1, \ldots, \mathbf{s}_K$ are presented over K trials yielding responses $r_1, \ldots, r_K$ with independent and identically distributed noise as in Equation 1, the maximum likelihood estimate (MLE, see **Table 2** for a table of abbreviations) for the filter $\mathbf{w}$ is given by

$$\hat{\mathbf{w}} = \underset{\mathbf{w}}{\operatorname{argmax}} \prod_{k=1}^{K} p(r_k | \mathbf{s}_k, \mathbf{w}). \tag{2}$$

Since the noise model is Gaussian, the solution to Equation (2) is simply the ordinary least squares solution (Bishop, 2006)

$$\hat{\mathbf{w}} = (\mathbf{S}^\top \mathbf{S})^{-1} \mathbf{S}^\top \mathbf{r} \tag{3}$$

where $\mathbf{S} = (\mathbf{s}_1, \ldots, \mathbf{s}_K)^\top$ is the stimulus design matrix and $\mathbf{r} = (r_1, \ldots, r_K)^\top$ is the vector of neuron responses.

A common interpretation of the estimator in Equation (3) is in terms of the spike-triggered average (STA) of the stimulus, which is the filter obtained by averaging over the stimuli that elicited a response,

$$\hat{\mathbf{w}}_{\mathrm{STA}} = \frac{1}{N} \mathbf{S}^\top \mathbf{r} \tag{4}$$

where N is the total number of spikes. When the stimulus ensemble follows a multivariate Gaussian with independent dimensions (and is therefore not biased toward any particular region of the feature space) the STA is the optimal filter (Chichilnisky, 2001; Dayan and Abbott, 2001; Simoncelli et al., 2004) and is proportional to the MLE. In general, the MLE pre-multiplies the STA by the inverse of the autocorrelation matrix $\mathbf{S}^\top \mathbf{S}$ of the stimulus ensemble to correct for bias in the presented stimuli, and thus corresponds to a whitened STA. Further discussion of the STA and its connection to the MLE can be found in Simoncelli et al. (2004) and Meyer et al. (2017).

2.2. The Linear-Nonlinear-Poisson Model

While a linear model can recover basic receptive field structure, it fails to capture the nonlinear changes in firing rate observed in electrophysiological recordings in cortex. In addition, the assumption of Gaussian noise leads to continuous (and possibly negative) estimates of spike counts. The linear-nonlinear-Poisson (LNP) model addresses these shortcomings by equipping the

FIGURE 1 | Generalized linear encoding models. **(A)** Basic linear-Gaussian model. A neuron's response is modeled as a linear combination of the stimulus components with additive Gaussian noise. In this example, the stimulus filter represents a two-dimensional visual receptive field. **(B)** The LNP model extends the linear-Gaussian model with a static nonlinearity applied to the filtered stimulus, together with Poisson observations to directly model spike counts. **(C)** Multi-neuron GLM encoding model used in Pillow et al. (2008). **(D)** Center (red) and surround (blue) components of temporal (left) and spatial (right) stimulus filters for the GLM fitted to an example ON retinal ganglion cell. **(E)** Same as **(D)**, but for an example OFF cell. In contrast to the ON cell, the OFF cell has an inhibitory effect on spiking. **(F)** Schematic of Bayesian decoding process. The encoding model $p(\mathbf{r}|\mathbf{s})$ is first fit by maximum likelihood. Then stimuli are decoded as the mean of the posterior stimulus distribution $p(\mathbf{s}|\mathbf{r})$ obtained by Bayes rule. **(G)** Performance of various decoders. Decoding using a GLM with interneuronal coupling filters (full model) substantially increases performance over models that do not account for interaction effects (linear, Poisson, and uncoupled models). **(C–G)** Adapted with permission from Pillow et al. (2008).

generative model with a static nonlinearity following the linear filtering, and a Poisson noise model to directly model the number of spikes generated within a fixed time-window (**Figure 1B**) (Chichilnisky, 2001). Let $t = 1, \ldots, T$ index over time bins. The LNP model assumes spikes follow an inhomogeneous Poisson process with time-varying firing rate $\lambda(t)$,

$$\lambda(t) = g(\mathbf{w}^\top \mathbf{s}(t)), \qquad r(t) \sim \text{Pois}(\lambda(t)) \qquad (5)$$

where g is a nonlinear activation function. While this nonlinearity can be estimated nonparametrically for each neuron (Simoncelli et al., 2004), it is often chosen to be $g(x) = \exp(x)$ as this ensures a non-negative intensity λ and tractable model fitting. Note that the specified firing rate $\lambda(t)$ will depend on the width Δ of the time bins or imaging rate, but for clarity here and for the remainder of the paper we omit explicit dependence of $\lambda(t)$ on Δ.

Assuming $g(x) = \exp(x)$ and that the responses $r(t)$ are count data, the MLE for the LNP model is the solution

$$\hat{\mathbf{w}} = \underset{\mathbf{w}}{\text{argmax}} \prod_{t=1}^{T} p(r(t)|\mathbf{s}(t), \mathbf{w}) = \underset{\mathbf{w}}{\text{argmax}} \sum_{t=1}^{T} \left(r(t) \ln \lambda(t) - \lambda(t) \right)$$

$$(6)$$

where the second equality follows by substituting the Poisson mass function and taking logarithms. The LNP model can be fit by standard gradient-based optimization methods since the intensity function $\lambda(t)$ is differentiable with respect to the filter parameters $\mathbf{w}$ and the log-likelihood function is concave (Paninski, 2004).

Regularization is a commonly used technique in machine learning for preventing a model from overfitting the training data. When maximizing the log-likelihood function for the LNP model with regularization, one penalizes the filter components whenever they deviate from zero

$$\hat{\mathbf{w}} = \underset{\mathbf{w}}{\text{argmax}} \sum_{t=1}^{T} (r(t) \ln \lambda(t) - \lambda(t)) - \eta ||\mathbf{w}||_p \qquad (7)$$

where $|| \cdot ||_p$ denotes the L_p norm and $\eta > 0$ is a penalty coefficient. Setting $p = 1$ or $p = 2$ corresponds, respectively, to LASSO and ridge regression (Friedman et al., 2001), encouraging a sparse filter $\mathbf{w}$. Maximizing the penalized log-likelihood is equivalent to performing posterior inference in a Bayesian regression model where $\mathbf{w}$ has a Laplacian (for $p = 1$) or Gaussian (for $p = 2$) prior (Wu et al., 2006). In many circumstances,

TABLE 1 | Table of notation.

Symbol	Parameter
r	Response
s	Stimulus
w	Stimulus filter
S	Stimulus design matrix
λ	Intensity, mean
$\epsilon, \xi, \xi_{h_\ell}$	Gaussian noise variable
σ, σ_S, ν	Noise standard deviation
$\boldsymbol{\mu}, \boldsymbol{\mu}_S$	Baseline activity, mean
h	Spike history vector
$\mathbf{J}_{ij}$	Spike history coupling filter
x	Latent factor
$\boldsymbol{\Psi}_S$	Diagonal variance matrix
$\boldsymbol{\Lambda}, \boldsymbol{\Lambda}_S$	Factor loading matrix
$\boldsymbol{\Sigma}_S$	Covariance matrix
τ	Gaussian process timescale
n	Spike count
c	Calcium concentration
γ	Autoregressive coefficient
α	Fluorescence scale
β	Baseline fluorescence level
f	Fluorescence level
p	Spike probability
h_ℓ	Refractory term
k_d	Dissociation constant
$\mathcal{N}$	Gaussian distribution
Pois	Poisson distribution
Bern	Bernoulli distribution

A subscripted s indicates stimulus-specific parameters.

TABLE 2 | Table of abbreviations.

Abbreviation	Meaning
MLE	Maximum likelihood estimate
STA	Spike-triggered average
LNP	Linear-nonlinear-Poisson
GLM	Generalized linear model
FA	Factor analysis
EM	Expectation maximization
GP	Gaussian process
GPFA	Gaussian process factor analysis
PLDS	Poisson linear dynamical system

such as when the data exhibits high noise levels, the ordinary (unpenalized) MLE cannot recover realistic receptive fields and needs to be constrained by regularization or priors (Sahani and Linden, 2003). Such Bayesian methods become highly effective in regimes of high noise, and a number of Bayesian extensions of receptive field inference invoke more subtle machine learning methods. For example, automatic relevance determination (Sahani and Linden, 2003) places a Gaussian prior on each element $\mathbf{w}_i$ of the filter and iteratively updates the prior variance until the filter components corresponding to irrelevant stimulus features effectively vanish from the model. Automatic locality determination, on the other hand, involves constructing receptive field priors encoding the information that receptive fields tend to be localized in space, time relative to the stimulus, and spatiotemporal frequency (Park and Pillow, 2011).

2.3. Extensions of the LNP Model

The LNP model is a special case of a generalized linear model (GLM): a class of encoding models that generalize the simple linear-Gaussian model to models that follow linear filtering with a static nonlinearity and any noise model from the exponential family. While there is in general no probability mass function for a multivariate extension of the Poisson distribution, the GLM framework allows one to incorporate interaction effects between different neurons, thereby allowing statistical models for single

neurons to be used for entire populations. The LNP model is extended by the addition of spike-history filters $\mathbf{J}_{ij}$ for all pairs of neurons i and j, intended to capture refractory effects for individual neurons (i.e., when $i = j$) and interaction effects between neurons ($i \neq j$), giving

$$\lambda_i(t) = \exp\left(\mathbf{w}_i^\top \mathbf{s}(t) + \sum_{j=1}^{N} \mathbf{J}_{ij}^\top \mathbf{h}_j(t)\right), \qquad r_i(t) \sim \text{Pois}(\lambda_i(t))$$

(8)

where $\mathbf{w}_i$ is the stimulus filter for neuron i, $\mathbf{h}_j(t) = (r_j(t - 1), \ldots, r_j(t - \tau))^\top$ is a vector of neuron j's spike history, and τ determines the length of the spike history window. The addition of the coupling filters allows the GLM to model the correlation structure within a population of neurons, as opposed to a model consisting of independent LNP neurons. Note, however, that the GLM is only well defined for coupling filters that act on the recent spike history of other neurons within the population, and cannot model correlations that arise from coactivity with zero time-lag (Macke et al., 2011). This motivates the use of latent variable models (see below), where simultaneous correlations arise among neurons whose activity is concurrently modulated by a shared factor.

Nonetheless, the GLM has been successfully applied to many data sets (Pillow et al., 2005, 2008; Park et al., 2014). Notably, Pillow et al. (2008) applied the GLM to a population of retinal ganglion cells from the fly (**Figures 1C–E**), obtained a complete characterization of the network's spatiotemporal correlation structure, and showed how incorporating these correlations yields a ∼20% increase in estimated information about the presented visual scene (**Figures 1F,G**).

3. LATENT FACTOR MODELS

3.1. Encoding With Factor Analysers

A frequent observation when recording population responses to the repeated presentation of identical stimuli is that variability tends to be correlated among groups of neurons. Such correlated variability (also known as shared variability or noise correlations) can substantially impact the efficacy of a neural code depending on the particular correlation structure (Abbott and Dayan, 1999; Schneidman et al., 2006; Lin et al., 2015), and suggests that there may be factors present that comodulate the responses of groups

of neurons. Factor analysis (FA), a probabilistic generalization of principal components analysis, is a classical model for inferring the latent group structure that can give rise to correlated variability.

In a Gaussian coding scheme with independent neurons, a population response $\mathbf{r}$ to a fixed stimulus s has a probability density given by

$$p(\mathbf{r}|s) = \mathcal{N}(\mathbf{r}|\boldsymbol{\mu}_s, \sigma_s^2 \mathbf{I}_N) \tag{9}$$

where the vector $\boldsymbol{\mu}_s$ is the mean population response, σ_s^2 is a noise variance common to each neuron, and $\mathbf{I}_N$ is the $N \times N$ identity matrix. While this model is analytically tractable with closed-form expressions for $\boldsymbol{\mu}_s$ and σ_s, the diagonal covariance matrix means it fails to account for the correlation structure that may be present in the data. As shown in e.g., Pillow et al. (2008), this additional information can considerably influence decoding accuracy.

On the other hand, a Gaussian model with an unconstrained covariance matrix $\boldsymbol{\Sigma}_s$ yields a density of the form

$$p(\mathbf{r}|s) = \mathcal{N}(\mathbf{r}|\boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s), \tag{10}$$

which, in principle, could outperform the Gaussian version that uses an unrealistic assumption of independently acting neurons (Santhanam et al., 2009). However, the covariance matrix $\boldsymbol{\Sigma}_s$ has $(N^2 + N)/2$ parameters to be learned per stimulus, requiring an amount of data that is impractically large to obtain experimentally for large N.

FA is a more moderate approach that attempts to capture shared variability in population activity by specifying a tractable parameterization of the covariance matrix. For FA the covariance matrix is defined as $\boldsymbol{\Sigma}_s = \boldsymbol{\Lambda}_s \boldsymbol{\Lambda}_s^\top + \boldsymbol{\Psi}_s$, where $\boldsymbol{\Psi}_s \in \mathbb{R}^{N \times N}$ is a diagonal matrix, $\boldsymbol{\Lambda}_s \in \mathbb{R}^{N \times q}$ is a factor loading matrix (analogous to the component loading matrix in principal components analysis), and $q < N$ determines the rank of $\boldsymbol{\Lambda}_s \boldsymbol{\Lambda}_s^\top$. Hence the population response $\mathbf{r}$ is distributed as

$$p(\mathbf{r}|s) = \mathcal{N}(\mathbf{r}|\boldsymbol{\mu}_s, \boldsymbol{\Lambda}_s \boldsymbol{\Lambda}_s^\top + \boldsymbol{\Psi}_s). \tag{11}$$

This decomposes $\boldsymbol{\Sigma}_s$ into two matrices that capture separate aspects of the response variability: $\boldsymbol{\Lambda}_s \boldsymbol{\Lambda}_s^\top$ is a low-rank matrix that captures the variability that is shared across neurons, whereas the diagonal matrix $\boldsymbol{\Psi}_s$ captures variability private to each neuron (Churchland et al., 2010). A critical observation is that the FA covariance matrix only requires $(q + 1)N$ parameters, which is less than $(N^2 + N)/2$ whenever $q < (N-1)/2$. Since q is usually chosen to be small, the FA covariance matrix requires much fewer parameters to be learned from the data.

An equivalent formulation of FA models the population response to a stimulus s as the projection from a low dimensional space of latent factors into the N-dimensional population space. This low dimensionality constraint forces any variability that the latent factors account for to be shared across groups of neurons, which leads to a modular correlation structure in the population recording. The generative model for the population response $\mathbf{r}_s$ given a stimulus s is

$$\mathbf{r}_s = \boldsymbol{\Lambda}_s \mathbf{x}_s + \boldsymbol{\mu}_s + \boldsymbol{\epsilon}_s \tag{12}$$

$$\mathbf{x}_s \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_q) \tag{13}$$

$$\boldsymbol{\epsilon}_s \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Psi}_s), \tag{14}$$

where $\mathbf{x}_s \in \mathbb{R}^q$ denotes the vector of latent factors, which are assumed to be independent with a Gaussian prior. These factors are intended to reflect unobserved brain states and could be physiologically realized as, e.g., shared gain modulation by downstream circuits. Note that the formulation of FA in Equation (11) can be recovered from Equations (12–14) by marginalizing over the latent factors.

Maximum likelihood estimation of the FA parameters $\theta_s = (\boldsymbol{\mu}_s, \boldsymbol{\Psi}_s, \boldsymbol{\Lambda}_s, \sigma_s)$ is complicated by the presence of latent variables $\mathbf{x}$, as the MLE $\hat{\theta}_s$ depends on an estimated $\hat{\mathbf{x}}$, and vice versa. FA thus uses the Expectation Maximization (EM) algorithm, an iterative procedure for fitting latent variable models (Dempster et al., 1977; Ghahramani et al., 1996). One must also choose the dimensionality q of the latent space, typically with a standard model selection procedure such as a comparison of the cross-validated log-likelihood or with an information criterion (Schwarz, 1978).

The FA method was applied to rhesus monkeys with brain-computer interfaces implanted in area PMd (Santhanam et al., 2009). Monkeys were trained on reaching tasks and the authors attempted to infer the intended target from electrophysiological data using a decoder based on the FA encoding model. By fitting the factor analyser, the decoder inferred the latent factors that comodulated neurons' responses. Incorporating this information led to substantial improvements in decoding accuracy over decoders based on independent Gaussian and Poisson encoding models.
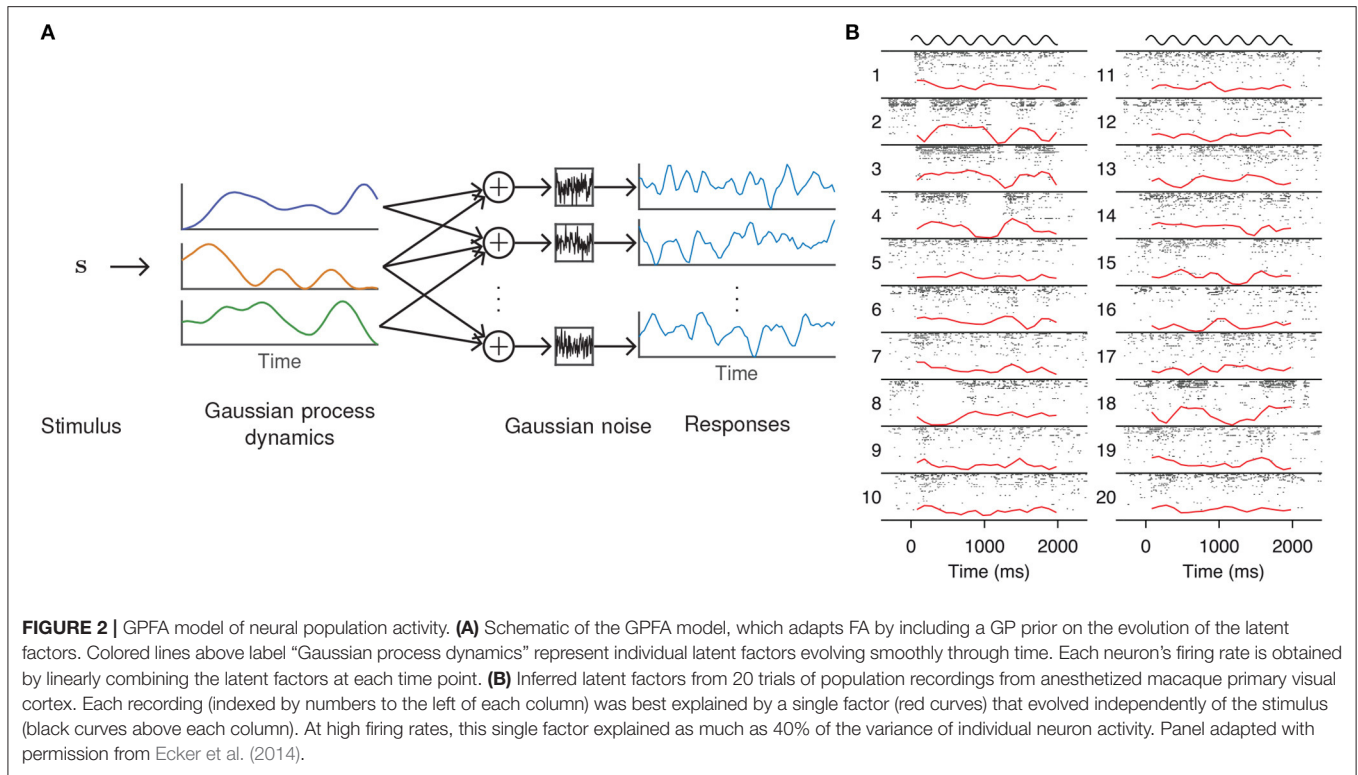
3.2. Gaussian Process Factor Analysis

The peristimulus time histogram averages spike trains over many trials to robustly estimate the aggregate effect of presenting a stimulus. Similarly, the FA encoding model is fit by pooling responses across trials to estimate the parameters θ_s. While this across-trial synthesis is necessary for fitting model parameters accurately, it will fail to reveal possibly important subtleties in neural activity within individual trials (Churchland et al., 2007, 2010; Afshar et al., 2011).

One way to adapt FA to single-trial analysis is to model the temporal evolution of the latent factors. A common technique in machine learning for enforcing temporal structure (or smoothness more generally) is Gaussian process (GP) regression, a Bayesian technique for nonparametric statistical modeling that places a GP prior on the latent variables (Williams and Rasmussen, 2006). The Gaussian process factor analysis (GPFA, **Figure 2A**) model (Yu et al., 2009) defines a GP for each dimension of the latent state $\ell = 1, \ldots, q$, which, in the case of discretely indexed time, reduces to a collection of multivariate Gaussians

$$\mathbf{x}^{(\ell)} \sim \mathcal{N}(\mathbf{0}, \mathbf{K}). \tag{15}$$

Here each $\mathbf{x}^{(\ell)} = (\mathbf{x}^{(\ell)}(1), \ldots, \mathbf{x}^{(\ell)}(T))^\top$. Elements of the covariance matrix $\mathbf{K}$ are typically determined by the squared

FIGURE 2 | GPFA model of neural population activity. **(A)** Schematic of the GPFA model, which adapts FA by including a GP prior on the evolution of the latent factors. Colored lines above label "Gaussian process dynamics" represent individual latent factors evolving smoothly through time. Each neuron's firing rate is obtained by linearly combining the latent factors at each time point. **(B)** Inferred latent factors from 20 trials of population recordings from anesthetized macaque primary visual cortex. Each recording (indexed by numbers to the left of each column) was best explained by a single factor (red curves) that evolved independently of the stimulus (black curves above each column). At high firing rates, this single factor explained as much as 40% of the variance of individual neuron activity. Panel adapted with permission from Ecker et al. (2014).

exponential kernel for encouraging smoothness

$$\mathbf{K}_{t_1, t_2} = \sigma_f^2 \exp\left(-\frac{(t_1 - t_2)^2}{2\tau^2}\right) + \sigma_n^2 \delta(t_1, t_2) \qquad (16)$$

where δ is the Kronecker delta function and σ_f and σ_n are parameters controlling the variance of the GP. The observed responses are then modeled as in FA,

$$\mathbf{r}(t) = \mathbf{\Lambda}\mathbf{x}(t) + \boldsymbol{\mu} + \boldsymbol{\epsilon}(t) \qquad (17)$$

$$\boldsymbol{\epsilon}(t) \sim \mathcal{N}(\mathbf{0}, \mathbf{\Psi}) \qquad (18)$$

where $\mathbf{x}(t)$ is the latent state at time t, $\mathbf{\Lambda}$ is the factor loading matrix, and $\boldsymbol{\mu}$ is a baseline activity level. GPFA can be viewed as a sequence of factor analysers (one for each time point) whose dimensions are linked together by smooth GPs. Note that while we have specified a single GP timescale τ, one can also assign distinct timescales τ_i to each dimension at the cost of an increase in computational overhead.

An advantage of GPFA is that the posterior over latent states $\mathbf{x}^{(\ell)}$ can be written down analytically because both the prior and likelihood are Gaussian, which form a conjugate pair (Bishop, 2006). This naturally leads to model fitting with the EM algorithm, where the updates for the parameter estimates are analogous to EM for FA (Ghahramani et al., 1996; Yu et al., 2009). Other examples of GP-based latent factor models are given in Nam (2015), Zhao and Park (2017), and Wu et al. (2017).

In a study of opioid anesthesia in macaque primary visual cortex, Ecker et al. (2014) used GPFA to investigate stimulus-driven patterns of population activity. The fitted model possessed

a single latent dimension that unmasked spontaneous transitions between periods of inactivity and highly elevated activity (**Figure 2B**). This single factor explained the observed increase in noise correlations and accounted for 40% of the variance of individual neuron firing rates. The extracted latent factors spanned a range of timescales, with some data best described by a latent factor whose strength changed slowly, on the order of several minutes. Similar up and down states had previously been seen only with non-opioid anesthetics.

3.3. The Poisson Linear Dynamical System

An alternative approach for latent trajectory modeling is to estimate the underlying linear dynamics of the latent state (Macke et al., 2011; Churchland et al., 2012; Pandarinath et al., 2018a). While the classical Kalman filter is the most thoroughly developed method for estimating the transition matrix in a linear dynamical system, a more appropriate generative model for neurons is the Poisson linear dynamical system (PLDS, **Figure 3A**) (Macke et al., 2011), which substitutes Poisson observations for the Gaussian emissions in the Kalman filter to directly model observed spike counts. The latent state $\mathbf{x}_k(t) \in \mathbb{R}^q$ on trial k at time bin t follows linear Markovian dynamics

$$\mathbf{x}_k(t+1) = \mathbf{A}\mathbf{x}_k(t) + \mathbf{b}(t) + \boldsymbol{\epsilon}_k(t+1) \qquad (19)$$

$$\mathbf{x}_k(1) \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_1) \qquad (20)$$

$$\boldsymbol{\epsilon}_k(t) \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}) \qquad (21)$$

where $\mathbf{A}$ is the dynamics matrix, $\mathbf{Q}$ is the noise covariance for the latent linear dynamics, and $\mathbf{Q}_1$ is the covariance of the initial
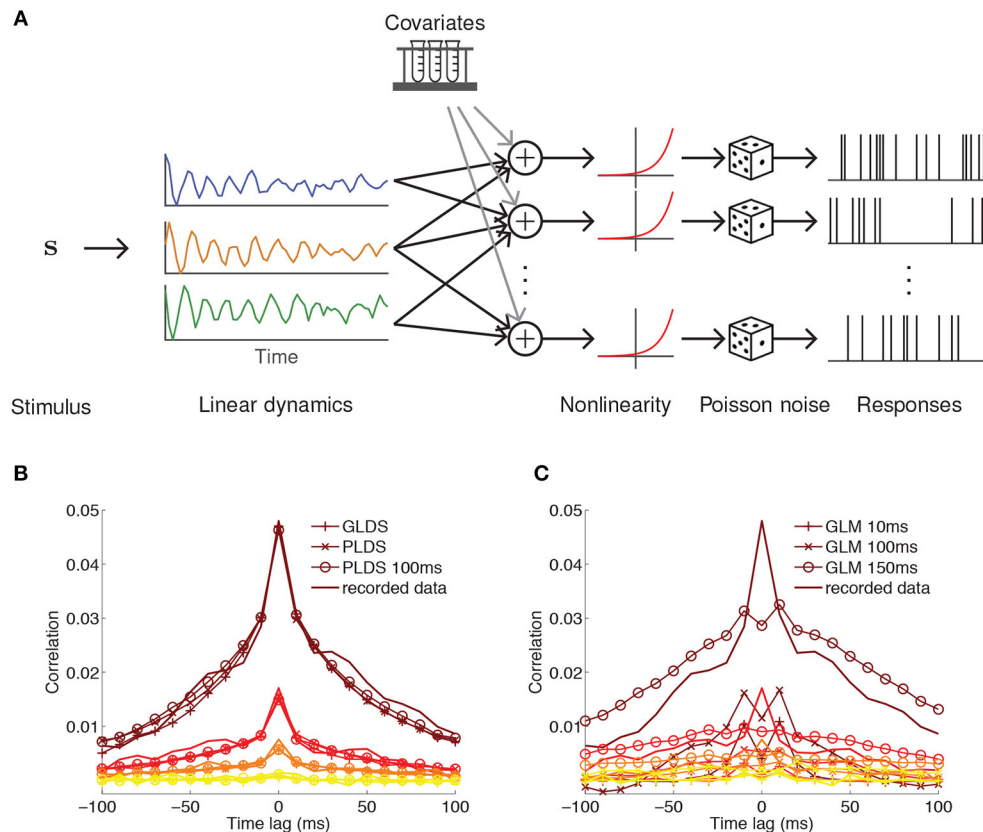
FIGURE 3 | PLDS model of neural population activity. **(A)** Schematic of the PLDS model. Analogous to GPFA, the PLDS model places a linear dynamical system prior over the latent factors. The activity of the factors is combined linearly, rectified by a nonlinearity, and determines Poisson spiking intensity for each neuron. Experimental parameters, spike history, and gain variables are often incorporated as covariates in the linear combination stage. **(B)** Averaged cross-correlations for latent dynamical system models with Gaussian observations (GLDS), Poisson observations (PLDS), and Poisson observations with spike history dependence (PLDS 100 ms). Groups of colored lines represent the average cross-correlation for the most correlated group of neurons (top group, brown) to the least correlated group (bottom group, yellow). Latent dynamical systems models have cross-correlations that align closely with the recorded data. **(C)** Same as **(B)**, but for GLMs with spike history filters of varying duration. Misalignment between the cross-correlations obtained from the model and the recorded data indicate that GLMs struggle to account for correlations at short time lags, in contrast to latent factor models where they arise naturally. **(B,C)** Adapted from Macke et al. (2011).

state. The latent dynamics are driven by a variable $\mathbf{b}(t)$ that captures stimulus-specific effects. Note that the PLDS model is formulated with explicit dependence on the trial index k, so that $\mathbf{b}(t)$ accounts for stimulus effects that are trial-independent. Similar to the LNP model, the observed spike responses on trial k then follow a Poisson distribution with mean $\lambda_{i,k}(t)$ derived from the latent state. For neuron i this takes the form

$$\lambda_{i,k}(t) = g(\mathbf{\Lambda}_{(i)}\mathbf{x}_k(t) + \boldsymbol{\mu}_i), \qquad r_{i,k}(t) \sim \text{Pois}(\lambda_{i,k}(t)). \quad (22)$$

Here the latent state influences an individual neuron i according to a row $\mathbf{\Lambda}_{(i)}$ of the factor loading matrix $\mathbf{\Lambda}$, and the low dimensionality of the latent state leads to the correlated variability as in the discussion of FA. Common choices for the nonlinearity include $g(x) = \exp(x)$ (Macke et al., 2011) and $g(x) = \ln(1 + \exp(x))$ (Buesing et al., 2017). This model can be modified in various ways to suit the data. For example, the stimulus drive term $\mathbf{b}$ in Equation (19) can be moved within the nonlinearity in Equation (22), so that the latent dynamics are decoupled from the stimulus and only

reflect changes internal to the brain. The intensity can be further extended by adding terms for, e.g., multiplicative gain (Buesing et al., 2017) and spike history (Macke et al., 2011) to capture refractory effects. A major advantage of latent factor models is their ability to account for correlations within short time intervals (**Figure 3B**), which GLMs struggle to match (**Figure 3C**).

The PLDS model is fit using a modified EM algorithm, which requires computing the posterior over the latent variables. Due to the Poisson observation model an analytic form of this posterior is unavailable. Typically one replaces the exact posterior by its Laplace approximation, which accelerates model fitting but violates some assumptions of the EM algorithm, resulting in an approximate inference framework (Macke et al., 2011).

An application of PLDS to multi-electrode recordings from songbird auditory cortex by Buesing et al. (2017) revealed that responses are modulated by shared variability with a single latent state, a similar result to Ecker et al. (2014). Buesing et al. histologically traced the locations of the recording sites and found a spatial gradient in the strength of the latent states. Shared variability was stronger (i.e., neurons were more strongly

coupled to the latent state) in deeper regions of auditory cortex. Interestingly, this strength was much weaker for certain stimulus classes than others, suggesting that deeper neurons selectively decouple from the latent state according to their stimulus preference. Other examples of dynamical systems-based latent factor models are given in Paninski et al. (2010), Buesing et al. (2012), Pfau et al. (2013), Semedo et al. (2014), Buesing et al. (2014), Archer et al. (2014), Kao et al. (2015), Gao et al. (2016), and Pandarinath et al. (2018b).

4. GENERATIVE MODELS FOR CALCIUM IMAGING DATA

4.1. Autoregressive Calcium Dynamics and Spike Deconvolution

The potential utility of large scale simultaneous neural recordings is constrained by our ability to make use of sophisticated techniques (such as latent factor methods) to analyse the data. While calcium imaging provides access to such large scale data, the models discussed so far assume that the data being analyzed is electrophysiological; i.e., that the neurons' responses are spike counts (for Poisson noise models) or firing rates (e.g., for Gaussian noise models). Their application to calcium imaging thus requires knowledge of how the optically recorded fluorescence signals are related to the underlying spiking activity. One approach to solving this problem involves constructing a generative statistical model where the spike counts are latent variables that are subsequently inferred from the fluorescence levels.

The presentation of a stimulus elicits a sequence of spikes across a population of neurons. For an individual neuron, we have assumed that the number of spikes within a time bin is sampled from a Poisson distribution with mean λ according to its particular receptive field. Each action potential is associated with a stereotypical rise and decay of the intracellular calcium concentration $c(t)$, usually modeled by an autoregressive process of order p (suppressing initial conditions for clarity) (Vogelstein et al., 2009),

$$c(t) = \sum_{i=1}^{p} \gamma_i c(t-i) + n(t), \qquad n(t) \sim \text{Pois}(\lambda) \quad (23)$$

where the Poisson-distributed random variable $n(t)$ models the generation of spikes within a time bin and $\gamma_1, \ldots, \gamma_p$ are the autoregressive coefficients that govern the rise and decay of the fluorescence levels. The observed fluorescence signal $f(t)$ is then obtained by a linear transformation of the calcium levels with additive noise,

$$f(t) = \alpha c(t) + \beta + \epsilon(t), \qquad \epsilon(t) \sim \mathcal{N}(0, \sigma^2) \quad (24)$$

where α sets the scale of the fluorescence signal and β accounts for a baseline fluorescence that may be unique to the imaging set-up or due to specific biophysical properties of individual neurons. The Gaussian noise model is intended to encompass variability due to, e.g., light scattering and shot noise (Delaney et al., 2018).

Note that this model does not set parameters for the scale or baseline of the calcium transient in Equation (23), as they are absorbed by α and β when the calcium is transformed to obtain the fluorescence (Vogelstein et al., 2009). An illustration of the generative model is given in **Figures 4A,B**.

For imaging systems where the rise time of the indicator is fast relative to the imaging rate a first-order autoregressive process is typically used, corresponding to an instantaneous rise and exponential decay of the calcium concentration. An autoregressive process of order 2 is used in situations where the rise time is slow relative to the imaging rate, in which case the calcium transient appears to approach its maximum amplitude gradually (Pnevmatikakis et al., 2016).

Models based on Equations (23, 24) have been used for spike train deconvolution (Vogelstein et al., 2009, 2010; Friedrich and Paninski, 2016; Pnevmatikakis et al., 2016). Let the vector $\theta = (\alpha, \beta, \lambda, \sigma, \{\gamma_i\}_{i=1}^{p})$ denote the model parameters, and let $\mathbf{f} = (f(1), \ldots, f(T))^{\top}$ and $\mathbf{n} = (n(1), \ldots, n(T))^{\top}$. Following Bayes' rule, the maximum *a posteriori* estimate for the spike train is

$$\hat{\mathbf{n}} = \underset{n(t) \in \mathbb{N}_0 \, \forall t}{\operatorname{argmax}} \, p(\mathbf{n}|\mathbf{f}, \theta) = \underset{n(t) \in \mathbb{N}_0 \, \forall t}{\operatorname{argmax}} \, p(\mathbf{f}|\mathbf{n}, \theta) p(\mathbf{n}|\theta) \quad (25)$$

where $\mathbb{N}_0$ is the set of non-negative integers. Given the spike sequence $\mathbf{n}$, the fluorescence levels $f(t)$ are independent and depend only on the calcium concentration $c(t)$, hence the likelihood factorizes as

$$p(\mathbf{f}|\mathbf{n}, \theta) = \prod_{t=1}^{T} p(f(t)|c(t), \theta) = \prod_{t=1}^{T} \mathcal{N}(f(t)|\alpha c(t) + \beta, \sigma^2). \quad (26)$$

Substituting Equation (26) into (25) and taking logarithms, the optimal sequence of spikes is then

$$\hat{\mathbf{n}} = \underset{n(t) \in \mathbb{N}_0 \, \forall t}{\operatorname{argmax}} \sum_{t=1}^{T} \left\{ -\frac{1}{2\sigma^2}(f(t) - \alpha c(t) - \beta)^2 + n(t) \ln \lambda - \ln(n(t)!) \right\}. \quad (27)$$

This is a difficult optimization problem because it requires searching through an infinite discrete space of spike trains. As noted in Vogelstein et al. (2010), even imposing an upper bound on the number of spikes within a frame yields an optimization problem with exponential computational complexity. One approach for overcoming this intractability involves approximating the Poisson distribution in Equation (25) by an exponential distribution, which leads to a concave objective function but with continuous estimates of $\hat{\mathbf{n}}$ (Vogelstein et al., 2010). This approximation also allows for a time-varying intensity function $\lambda(t)$, but does not explicitly model the transformation from stimulus to spiking intensity.

Runyan et al. (2017) applied a combination of the methods described in this review to study the timescales of population codes in cortex. 2-photon calcium imaging of auditory and posterior parietal cortices was performed while mice completed a sound localization task. The resulting fluorescence data
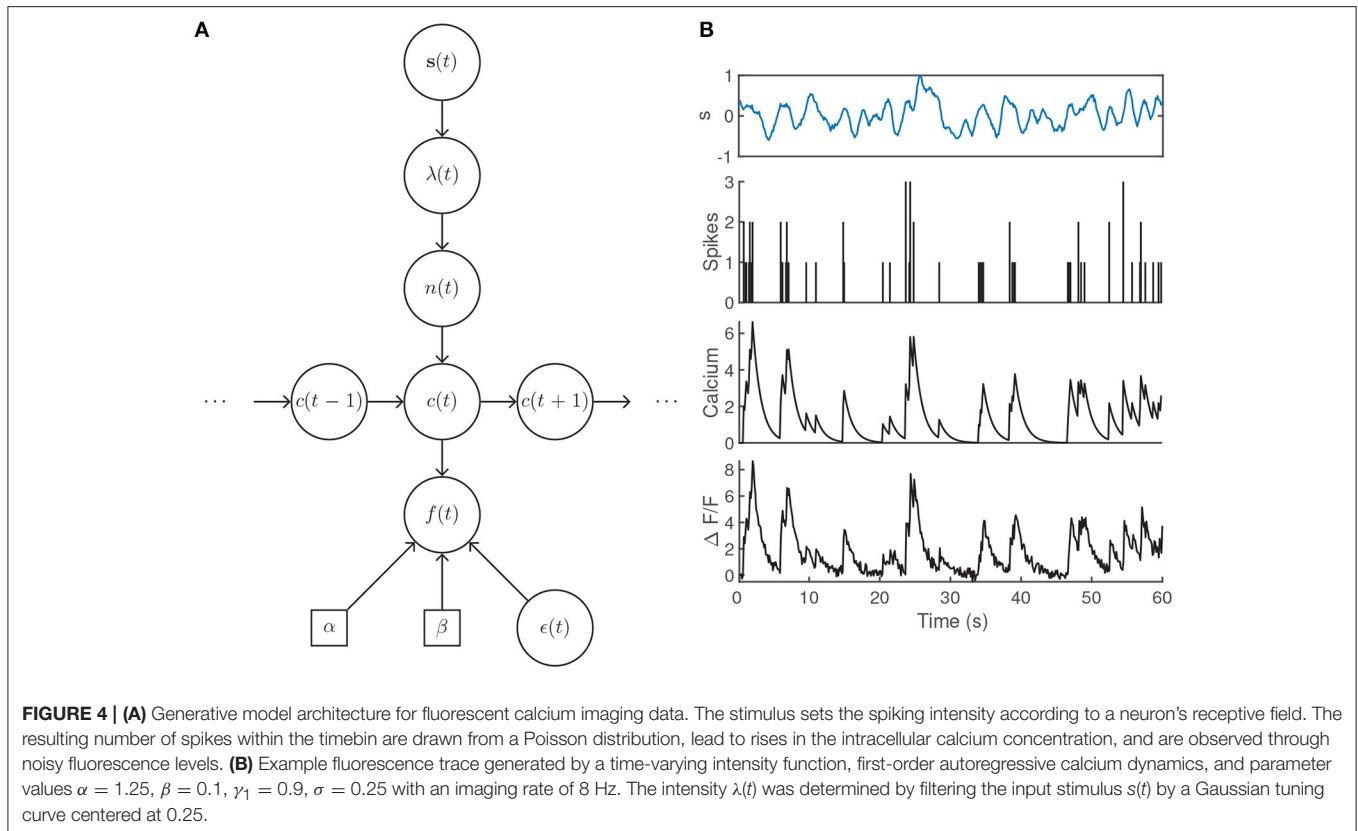
FIGURE 4 | (A) Generative model architecture for fluorescent calcium imaging data. The stimulus sets the spiking intensity according to a neuron's receptive field. The resulting number of spikes within the timebin are drawn from a Poisson distribution, lead to rises in the intracellular calcium concentration, and are observed through noisy fluorescence levels. **(B)** Example fluorescence trace generated by a time-varying intensity function, first-order autoregressive calcium dynamics, and parameter values $\alpha = 1.25$, $\beta = 0.1$, $\gamma_1 = 0.9$, $\sigma = 0.25$ with an imaging rate of 8 Hz. The intensity $\lambda(t)$ was determined by filtering the input stimulus $s(t)$ by a Gaussian tuning curve centered at 0.25.

was deconvolved according to the exponential-approximation approach described above to estimate firing rates (Vogelstein et al., 2010). They then fitted a GLM encoding model to populations from each cortical area that included coupling filters and various experimental and behavioral covariates. The fitted model was used in a decoding analysis that quantified the contribution of interneuronal coupling in the two cortical areas, and showed that stronger coupling was associated with population codes that had longer timescales. This provided evidence for a coding mechanism where tightly coupled populations of neurons prolonged the representation of stimuli through their sequential activation.

4.2. A Generalized Model for Calcium Dynamics

The calcium kinetics in Equation (23) are deterministic given the spike counts. In reality the concentration of calcium may be subject to many sources of variability, and analyses of some data sets may benefit from explicitly accounting for this noise. Vogelstein et al. (2009) modeled this by driving the calcium levels by both Bernoulli-distributed spikes and additive Gaussian noise,

$$c(t) = \gamma c(t-1) + n(t) + \xi(t) \tag{28}$$

$$n(t) \sim \text{Bern}(p(t)) \tag{29}$$

$$\xi(t) \sim \mathcal{N}(0, \nu^2) \tag{30}$$

where $\text{Bern}(p(t))$ is the Bernoulli distribution with time-dependent trial-success probability $p(t)$, $\gamma < 1$ is an autoregressive coefficient, and ν^2 is the calcium noise variance. A simplifying assumption in models based on Equations (23) and (24) is that spikes are generated independently of their spike history. However, the spike probability can be more generally modeled with a GLM (Vogelstein et al., 2009)

$$p(t) = 1 - \exp\left(-g\left(\mathbf{w}^\top \mathbf{s}(t) + \mathbf{J}^\top \mathbf{h}(t)\right)\right) \tag{31}$$

where g is a selected nonlinearity. Unlike the standard GLM structure of Equation 8, the spike history term here takes the form $\mathbf{h}(t) = (h_1(t), \ldots, h_L(t))^\top$, where each h_ℓ is an exponentially decaying refractory term that jumps following each spike

$$h_\ell(t) = \gamma_{h_\ell} h_\ell(t-1) + n(t) + \xi_{h_\ell}(t), \qquad \xi_{h_\ell}(t) \sim \mathcal{N}(0, \nu_{h_\ell}^2). \tag{32}$$

Finally, rather than a simple linear relationship between $f(t)$ and $c(t)$, Vogelstein et al. (2009) and Vogelstein et al. (2010) also consider saturating fluorescence levels using a nonlinear Hill function with dissociation constant k_d

$$f(t) = \alpha \frac{c(t)}{c(t) + k_d} + \beta + \epsilon(t), \qquad \epsilon(t) \sim \mathcal{N}(0, \sigma^2). \tag{33}$$

Importantly, saturation of the fluorescence signal causes the spike-triggered fluorescence transients to become progressively

smaller during a train of action potentials, and failure to account for this detail may limit the accuracy of spike deconvolution algorithms. The model defined by Equations (28–33) is fit using a sequential Monte Carlo method (Vogelstein et al., 2009). By including explicit stimulus and spike history filters, Vogelstein et al. (2009) could accurately infer spike times from fluorescence data with temporal superresolution; i.e., could identify when within an imaging frame each spike occurs. Some other example methods for spike deconvolution are based on compressed sensing (Pnevmatikakis and Paninski, 2013), fully Bayesian inference (Pnevmatikakis et al., 2013), and variational autoencoders (Speiser et al., 2017).

5. DISCUSSION

Probabilistic modeling provides a practical, interpretable, and theoretically grounded framework for probing how networks of neurons process information. Many of the statistical models discussed in this review are abstract mathematical descriptions of how stimuli are related to patterns of neural activity. Often the mathematical operations that define the models do not necessarily attempt to align with real biological functions or behavior. Rather, such models are intended to serve as tools to uncover interpretable patterns and relationships that may not be detectable by other approaches. On the other hand, there are cases where the goal is to infer biophysical variables, as in e.g., models for calcium imaging data or for the anatomical architecture of a neural circuit, and then greater care must be taken to constrain the model by relevant physiological data (Paninski et al., 2007; Real et al., 2017; Latimer et al., 2018).

Recent advances in statistical models of spike train data have focused on incorporating more general nonlinear transformations of the latent state, including the use of neural networks (Gao et al., 2016; Pandarinath et al., 2018b) and GPs (Wu et al., 2017). This is in contrast to e.g., the FA and GPFA encoding models, where the mean spiking intensity of a neuron is obtained by a simple linear transformation of the latent state. Bayesian methods, such as latent factor modeling, are a powerful way to incorporate prior knowledge when making inferences about the behavior of a system. While GPFA places a smoothness prior on the evolution of latent factors to encourage some degree of temporal structure, other methods place priors on, e.g. network structure for connectivity inference (Linderman S. et al., 2016) and the latent states of a hidden Markov model with Poisson observations (Linderman S. W. et al., 2016).

Although there has been a rapid expansion in the number of models for extracting receptive fields, interneuronal coupling strengths, and latent structure from multivariate electrophysiological recordings, similar models for calcium imaging data are only beginning to emerge (Aitchison et al., 2017; Khan et al., 2018). A common approach for analysing calcium imaging data involves first deconvolving fluorescence traces and then fitting conventional models, but deconvolution methods only provide coarse estimates of firing rates. Spike trains obtained by highly optimized algorithms typically only agree with ground truth recordings with a correlation coefficient less than ~0.75, even with substantial training data, suggesting that there is an unavoidable loss of information associated with

spike deconvolution (Pnevmatikakis et al., 2016; Berens et al., 2018). An advantage of GPFA over earlier methods for estimating trajectories of population activity is that it condenses the two stages of dimensionality reduction and smoothing into a single stage of posterior inference. Similarly, probabilistic analysis of calcium imaging data can have the two stages of deconvolution and model fitting merged into a single step by marginalizing over possible spike trains (Ganmor et al., 2016), mitigating some of the information loss accompanied by deconvolution. Neural encoding models for calcium imaging data that avoid an explicit intermediate step of spike inference are likely to be an important future development in this area (Aitchison et al., 2017).

Many studies consider the amplitude of an evoked calcium transient as a measure of a neuron's response. This has been widely used in zebrafish larvae, for which there has been significant interest in recent years. For example, 2-photon calcium imaging of the zebrafish optic tectum has led to new insights into the circuit architecture determining selectivity to size, location, and direction of motion (Del Bene et al., 2010; Gabriel et al., 2012; Grama and Engert, 2012; Nikolaou et al., 2012; Lowe et al., 2013; Preuss et al., 2014; Avitan et al., 2016; Abbas et al., 2017), and light-sheet microscopy has allowed for the creation of brain-wide functional circuit models for motor behavior driven by vision (Naumann et al., 2016) and thermosensation (Haesemeyer et al., 2018). Similar studies in the future provide further opportunities for model-based analyses.

The techniques described in this review were developed for spike train or calcium imaging data, but some approaches are broadly applicable across systems neuroscience. For instance, suitably adapted latent factor models have been successfully applied to recordings of the local field potential, where it was found that the activity of particular latent factors could discriminate vulnerability to stress-induced behavioral dysfunction in mouse models of major depressive disorder (Gallagher et al., 2017; Hultman et al., 2018).

As the scale of multi-neuron data continues to grow, the creation of new models and their associated fitting algorithms may be spurred more by efficiency and scalability considerations than the level of statistical detail they are able to extract from experimental data (Zoltowski and Pillow, 2018). In some cases the computational issues associated with neural data analysis are more profound than simply needing a larger computer cluster. Neuropixel electrode arrays (Jun et al., 2017), for example, are capable of recording from hundreds of channels simultaneously, and may put inference algorithms under strain if computational efficiency is not sufficiently addressed. When combined with fluorescent sensors of neural activity, optogenetic photostimulation grants the ability to manipulate neural circuits in real time, and models are now beginning to explicitly integrate the effect of photostimulation on calcium transients (Aitchison et al., 2017). Moreover, genetically encoded voltage indicators operate on a timescale of tens of milliseconds (Knöpfel et al., 2015), overcoming one of the principal drawbacks of calcium imaging; namely, the slow binding kinetics of the indicator relative to the timescale of action potential generation. Combining these emerging technologies with models designed to capture their associated generative processes thus promises

to greatly improve our capacity to uncover how patterns of neural activity represent and process features of the external world.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

REFERENCES

Abbas, F., Triplett, M. A., Goodhill, G. J., and Meyer, M. P. (2017). A three-layer network model of direction selective circuits in the optic tectum. *Front. Neural Circuits* 11:88. doi: 10.3389/fncir.2017.00088

Abbott, L. F., and Dayan, P. (1999). The effect of correlated variability on the accuracy of a population code. *Neural Comput.* 11, 91–101. doi: 10.1162/089976699300016827

Abdelfattah, A. S., Kawashima, T., Singh, A., Novak, O., Liu, H., Shuai, Y., et al. (2018). Bright and photostable chemigenetic indicators for extended *in vivo* voltage imaging. *bioRxiv:436840*. doi: 10.1101/436840

Afshar, A., Santhanam, G., Yu, B. M., Ryu, S. I., Sahani, M., and Shenoy, K. V. (2011). Single-trial neural correlates of arm movement preparation. *Neuron* 71, 555–564. doi: 10.1016/j.neuron.2011.05.047

Ahrens, M. B., Orger, M. B., Robson, D. N., Li, J. M., and Keller, P. J. (2013). Whole-brain functional imaging at cellular resolution using light-sheet microscopy. *Nat. Methods* 10, 413–420. doi: 10.1038/nmeth.2434

Aitchison, L., Russell, L., Packer, A. M., Yan, J., Castonguay, P., Häusser, M., et al. (2017). "Model-based Bayesian inference of neural activity and connectivity from all-optical interrogation of a neural circuit," in *Advances in Neural Information Processing Systems*, eds I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Curran Associates, Inc.), 3489–3498.

Archer, E. W., Koster, U., Pillow, J. W., and Macke, J. H. (2014). "Low dimensional models of neural population activity in sensory cortical circuits," in *Advances in Neural Information Processing Systems*, eds Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger (Curran Associates, Inc.), 343–351.

Avitan, L., Pujic, Z., Hughes, N. J., Scott, E. K., and Goodhill, G. J. (2016). Limitations of neural map topography for decoding spatial information. *J. Neurosci.* 36, 5385–5396. doi: 10.1523/JNEUROSCI.0385-16.2016

Berens, P., Freeman, J., Deneux, T., Chenkov, N., McColgan, T., Speiser, A., et al. (2018). Community-based benchmarking improves spike rate inference from two-photon calcium imaging data. *PLoS Comput. Biol.* 14:e1006157. doi: 10.1371/journal.pcbi.1006157

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Berlin; Heidelberg: Springer-Verlag.

Buesing, L., Calabrese, A., Cunningham, J., Woolley, S., and Paninski, L. (2017). A statistical model of shared variability in the songbird auditory system. *bioRxiv:113670*. doi: 10.1101/113670

Buesing, L., Machado, T. A., Cunningham, J. P., and Paninski, L. (2014). "Clustered factor analysis of multineuronal spike data," in *Advances in Neural Information Processing Systems*, eds Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger (Curran Associates, Inc.), 3500–3508.

Buesing, L., Macke, J. H., and Sahani, M. (2012). Learning stable, regularised latent models of neural population dynamics. *Network* 23, 24–47. doi: 10.3109/0954898X.2012.677095

Chen, T.-W., Wardill, T. J., Sun, Y., Pulver, S. R., Renninger, S. L., Baohan, A., et al. (2013). Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature* 499, 295–300. doi: 10.1038/nature12354

Chen, X., Mu, Y., Hu, Y., Kuan, A. T., Nikitchenko, M., Randlett, O., et al. (2018). Brain-wide organization of neuronal activity and convergent sensorimotor transformations in larval zebrafish. *Neuron* 100, 876–890.e5. doi: 10.1016/j.neuron.2018.09.042

Chichilnisky, E. (2001). A simple white noise analysis of neuronal light responses. *Network* 12, 199–213. doi: 10.1080/713663221

Churchland, M. M., Cunningham, J. P., Kaufman, M. T., Foster, J. D., Nuyujukian, P., Ryu, S. I., et al. (2012). Neural population dynamics during reaching. *Nature* 487, 51–56. doi: 10.1038/nature11129

Churchland, M. M., Yu, B. M., Cunningham, J. P., Sugrue, L. P., Cohen, M. R., Corrado, G. S., et al. (2010). Stimulus onset quenches neural variability: a widespread cortical phenomenon. *Nat. Neurosci.* 13, 369–378. doi: 10.1038/nn.2501

Churchland, M. M., Yu, B. M., Sahani, M., and Shenoy, K. V. (2007). Techniques for extracting single-trial activity patterns from large-scale neural recordings. *Curr. Opin. Neurobiol.* 17, 609–618. doi: 10.1016/j.conb.2007.11.001

Cunningham, J. P., and Yu, B. M. (2014). Dimensionality reduction for large-scale neural recordings. *Nat. Neurosci.* 17, 1500–1509. doi: 10.1038/nn.3776

Dayan, P., and Abbott, L. F. (2001). *Theoretical Neuroscience*, vol. 806. Cambridge, MA: MIT Press.

Del Bene, F., Wyart, C., Robles, E., Tran, A., Looger, L., Scott, E. K., et al. (2010). Filtering of visual information in the tectum by an identified neural circuit. *Science* 330, 669–673. doi: 10.1126/science.1192949

Delaney, T., O'Donnell, C., and Ashby, M. (2018). *How Do Calcium Indicator Properties Affect Spike Inference?* Denver, CO: Cosyne Abstracts.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. Ser. B (Methodol.)* 39, 1–38.

Ecker, A. S., Berens, P., Cotton, R. J., Subramaniyan, M., Denfield, G. H., Cadwell, C. R., et al. (2014). State dependence of noise correlations in macaque primary visual cortex. *Neuron* 82, 235–248. doi: 10.1016/j.neuron.2014.02.006

Friedman, J., Hastie, T., and Tibshirani, R. (2001). *The Elements of Statistical Learning.* New York, NY: Springer-Verlag.

Friedrich, J., and Paninski, L. (2016). "Fast active set methods for online spike inference from calcium imaging," in *Advances in Neural Information Processing Systems*, eds D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett (Curran Associates, Inc.), 1984–1992.

Gabriel, J. P., Trivedi, C. A., Maurer, C. M., Ryu, S., and Bollmann, J. H. (2012). Layer-specific targeting of direction-selective neurons in the zebrafish optic tectum. *Neuron* 76, 1147–1160. doi: 10.1016/j.neuron.2012.12.003

Gallagher, N., Ulrich, K. R., Talbot, A., Dzirasa, K., Carin, L., and Carlson, D. E. (2017). "Cross-spectral factor analysis," in *Advances in Neural Information Processing Systems*, eds I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Curran Associates, Inc.), 6842–6852.

Ganmor, E., Krumin, M., Rossi, L. F., Carandini, M., and Simoncelli, E. P. (2016). Direct estimation of firing rates from calcium imaging data. arXiv: 1601.00364.

Gao, Y., Archer, E. W., Paninski, L., and Cunningham, J. P. (2016). "Linear dynamical neural population models through nonlinear embeddings," in *Advances in Neural Information Processing Systems*, eds D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett (Curran Associates, Inc.), 163–171.

Ghahramani, Z., and Hinton, G. E. (1996). *The EM Algorithm for Mixtures of Factor Analyzers*. Technical report, Technical Report CRG-TR-96-1, University of Toronto.

Grama, A., and Engert, F. (2012). Direction selectivity in the larval zebrafish tectum is mediated by asymmetric inhibition. *Front. Neural Circuits* 6:59. doi: 10.3389/fncir.2012.00059

Haesemeyer, M., Robson, D. N., Li, J. M., Schier, A. F., and Engert, F. (2018). A brain-wide circuit model of heat-evoked swimming behavior in larval zebrafish. *Neuron* 98, 817–831.e6. doi: 10.1016/j.neuron.2018.04.013

Hultman, R., Ulrich, K., Sachs, B. D., Blount, C., Carlson, D. E., Ndubuizu, N., et al. (2018). Brain-wide electrical spatiotemporal dynamics encode depression vulnerability. *Cell* 173, 166–180. doi: 10.1016/j.cell.2018.02.012

Jun, J. J., Steinmetz, N. A., Siegle, J. H., Denman, D. J., Bauza, M., Barbarits, B., et al. (2017). Fully integrated silicon probes for high-density recording of neural activity. *Nature* 551, 232–236. doi: 10.1038/nature24636

Kao, J. C., Nuyujukian, P., Ryu, S. I., Churchland, M. M., Cunningham, J. P., and Shenoy, K. V. (2015). Single-trial dynamics of motor cortex and their applications to brain-machine interfaces. *Nat. Commun.* 6:7759. doi: 10.1038/ncomms8759

Khan, A. G., Poort, J., Chadwick, A., Blot, A., Sahani, M., Mrsic-Flogel, T. D., et al. (2018). Distinct learning-induced changes in stimulus selectivity and interactions of GABAergic interneuron classes in visual cortex. *Nat. Neurosci.* 21, 851–859. doi: 10.1038/s41593-018-0143-z

Knöpfel, T., Gallero-Salas, Y., and Song, C. (2015). Genetically encoded voltage indicators for large scale cortical imaging come of age. *Curr. Opin. Chem. Biol.* 27, 75–83. doi: 10.1016/j.cbpa.2015.06.006

Latimer, K., Rieke, F., and Pillow, J. W. (2018). Inferring synaptic inputs from spikes with a conductance-based neural encoding model. bioRxiv 281089. doi: 10.1101/281089

Lin, I.-C., Okun, M., Carandini, M., and Harris, K. D. (2015). The nature of shared cortical variability. *Neuron* 87, 644–656. doi: 10.1016/j.neuron.2015.06.035

Linderman, S., Adams, R. P., and Pillow, J. W. (2016). "Bayesian latent structure discovery from multi-neuron recordings," in *Advances in Neural Information Processing Systems*, eds D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett (Curran Associates, Inc.), 2002–2010.

Linderman, S. W., Johnson, M. J., Wilson, M. A., and Chen, Z. (2016). A Bayesian nonparametric approach for uncovering rat hippocampal population codes during spatial navigation. *J. Neurosci. Methods* 263, 36–47. doi: 10.1016/j.jneumeth.2016.01.022

Lowe, A. S., Nikolaou, N., Hunter, P. R., Thompson, I. D., and Meyer, M. P. (2013). A systems-based dissection of retinal inputs to the zebrafish tectum reveals different rules for different functional classes during development. *J. Neurosci.* 33, 13946–13956. doi: 10.1523/JNEUROSCI.1866-13.2013

Macke, J. H., Buesing, L., Cunningham, J. P., Byron, M. Y., Shenoy, K. V., and Sahani, M. (2011). "Empirical models of spiking in neural populations," in *Advances in Neural Information Processing Systems*, eds J. Shawe-Taylor, R. S. Zemel, P. L. Bartlett, F. Pereira, and K. Q. Weinberger (Curran Associates, Inc.), 1350–1358.

Meyer, A. F., Williamson, R. S., Linden, J. F., and Sahani, M. (2017). Models of neuronal stimulus-response functions: elaboration, estimation, and evaluation. *Front. Syst. Neurosci.* 10:109. doi: 10.3389/fnsys.2016.00109

Nam, H. (2015). *Poisson Extension of Gaussian Process Factor Analysis for Modeling Spiking Neural Populations*. Master's thesis, Department of Neural Computation and Behaviour, Max Planck Institute for Biological Cybernetics, Tübingen.

Naumann, E. A., Fitzgerald, J. E., Dunn, T. W., Rihel, J., Sompolinsky, H., and Engert, F. (2016). From whole-brain data to functional circuit models: the zebrafish optomotor response. *Cell* 167, 947–960. doi: 10.1016/j.cell.2016.10.019

Nikolaou, N., Lowe, A. S., Walker, A. S., Abbas, F., Hunter, P. R., Thompson, I. D., et al. (2012). Parametric functional maps of visual inputs to the tectum. *Neuron* 76, 317–324. doi: 10.1016/j.neuron.2012.08.040

Pandarinath, C., Ames, K. C., Russo, A. A., Farshchian, A., Miller, L. E., Dyer, E. L., et al. (2018a). Latent factors and dynamics in motor cortex and their application to brain–machine interfaces. *J. Neurosci.* 38, 9390–9401. doi: 10.1523/JNEUROSCI.1669-18.2018

Pandarinath, C., O'Shea, D. J., Collins, J., Jozefowicz, R., Stavisky, S. D., Kao, J. C., et al. (2018b). Inferring single-trial neural population dynamics using sequential auto-encoders. *Nat. Methods* 15, 805–815. doi: 10.1038/s41592-018-0109-9

Paninski, L. (2004). Maximum likelihood estimation of cascade point-process neural encoding models. *Network* 15, 243–262. doi: 10.1088/0954-898X_15_4_002

Paninski, L., Ahmadian, Y., Ferreira, D. G., Koyama, S., Rahnama Rad, K., Vidne, M., et al. (2010). A new look at state-space models for neural data. *J. Comput. Neurosci.* 29, 107–126. doi: 10.1007/s10827-009-0179-x

Paninski, L., and Cunningham, J. P. (2018). Neural data science: accelerating the experiment-analysis-theory cycle in large-scale neuroscience. *Curr. Opin. Neurobiol.* 50, 232–241. doi: 10.1016/j.conb.2018.04.007

Paninski, L., Pillow, J., and Lewi, J. (2007). Statistical models for neural encoding, decoding, and optimal stimulus design. *Prog. Brain Res.* 165, 493–507. doi: 10.1016/S0079-6123(06)65031-0

Park, I. M., Meister, M. L., Huk, A. C., and Pillow, J. W. (2014). Encoding and decoding in parietal cortex during sensorimotor decision-making. *Nat. Neurosci.* 17, 1395–1403. doi: 10.1038/nn.3800

Park, M., and Pillow, J. W. (2011). Receptive field inference with localized priors. *PLoS Comput. Biol.* 7:e1002219. doi: 10.1371/journal.pcbi.1002219

Pfau, D., Pnevmatikakis, E. A., and Paninski, L. (2013). "Robust learning of lowdimensional dynamics from large neural ensembles," in *Advances in Neural Information Processing Systems*, eds C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger (Curran Associates, Inc.), 2391–2399.

Pillow, J. (2007). "Likelihood-based approaches to modeling the neural code," in *Bayesian Brain: Probabilistic Approaches to Neural Coding*, eds K. Doya, S. Ishii, A. Pouget, and R. P. N. Rao. (Cambridge, MA: MIT Press), 53–70.

Pillow, J. W., Paninski, L., Uzzell, V. J., Simoncelli, E. P., and Chichilnisky, E. (2005). Prediction and decoding of retinal ganglion cell responses with a probabilistic spiking model. *J. Neurosci.* 25, 11003–11013. doi: 10.1523/JNEUROSCI.3305-05.2005

Pillow, J. W., Shlens, J., Paninski, L., Sher, A., Litke, A. M., Chichilnisky, E., et al. (2008). Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature* 454, 995–999. doi: 10.1038/nature 07140

Pnevmatikakis, E. A., Merel, J., Pakman, A., and Paninski, L. (2013). "Bayesian spike inference from calcium imaging data," in *Signals, Systems and Computers, 2013 Asilomar Conference on* (Pacific Grove, CA: IEEE), 349–353.

Pnevmatikakis, E. A., and Paninski, L. (2013). "Sparse nonnegative deconvolution for compressive calcium imaging: algorithms and phase transitions," in *Advances in Neural Information Processing Systems*, eds C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger (Curran Associates, Inc.), 1250–1258.

Pnevmatikakis, E. A., Soudry, D., Gao, Y., Machado, T. A., Merel, J., Pfau, D., et al. (2016). Simultaneous denoising, deconvolution, and demixing of calcium imaging data. *Neuron* 89, 285–299. doi: 10.1016/j.neuron.2015. 11.037

Preuss, S. J., Trivedi, C. A., vom Berg-Maurer, C. M., Ryu, S., and Bollmann, J. H. (2014). Classification of object size in retinotectal microcircuits. *Curr. Biol.* 24, 2376–2385. doi: 10.1016/j.cub.2014.09.012

Quian Quiroga R., and Panzeri, S. (2009). Extracting information from neuronal populations: information theory and decoding approaches. *Nat. Rev. Neurosci.* 10, 173–185. doi: 10.1038/nrn2578

Real, E., Asari, H., Gollisch, T., and Meister, M. (2017). Neural circuit inference from function to structure. *Curr. Biol.* 27, 189–198. doi: 10.1016/j.cub.2016.11.040

Runyan, C. A., Piasini, E., Panzeri, S., and Harvey, C. D. (2017). Distinct timescales of population coding across cortex. *Nature* 548, 92–96. doi: 10.1038/nature23020

Sahani, M., and Linden, J. F. (2003). "Evidence optimization techniques for estimating stimulus-response functions," in *Advances In Neural Information Processing Systems*, eds S. Becker, S. Thrun, and K. Obermayer (MIT Press), 317–324.

Santhanam, G., Yu, B. M., Gilja, V., Ryu, S. I., Afshar, A., Sahani, M., et al. (2009). Factor-analysis methods for higher-performance neural prostheses. *J. Neurophysiol.* 102, 1315–1330. doi: 10.1152/jn.00097.2009

Schneidman, E., Berry II, M. J., Segev, R., and Bialek, W. (2006). Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature* 440, 1007–1012. doi: 10.1038/nature04701

Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Stat.* 6, 461–464. doi: 10.1214/aos/1176344136

Semedo, J., Zandvakili, A., Kohn, A., Machens, C. K., and Byron, M. Y. (2014). "Extracting latent structure from multiple interacting neural populations," in *Advances in Neural Information Processing Systems*, eds Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger (Curran Associates, Inc.), 2942–2950.

Simoncelli, E. P., Paninski, L., Pillow, J., and Schwartz, O. (2004). *Characterization of Neural Responses With Stochastic Stimuli*, Vol. 3. Cambridge, MA: MIT Press.

Speiser, A., Yan, J., Archer, E. W., Buesing, L., Turaga, S. C., and Macke, J. H. (2017). "Fast amortized inference of neural activity from calcium imaging data with variational autoencoders," in *Advances in Neural Information Processing Systems*, eds I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Curran Associates, Inc.), 4024–4034.

Vogelstein, J. T., Packer, A. M., Machado, T. A., Sippy, T., Babadi, B., Yuste, R., et al. (2010). Fast nonnegative deconvolution for spike train inference from population calcium imaging. *J. Neurophysiol.* 104, 3691–3704. doi: 10.1152/jn.01073.2009

Vogelstein, J. T., Watson, B. O., Packer, A. M., Yuste, R., Jedynak, B., and Paninski, L. (2009). Spike inference from calcium imaging using sequential Monte Carlo methods. *Biophys. J.* 97, 636–655. doi: 10.1016/j.bpj.2008.08.005

Williams, C. K., and Rasmussen, C. E. (2006). *Gaussian Processes for Machine Learning*. Cambridge, MA: MIT Press.

Wu, A., Roy, N. G., Keeley, S., and Pillow, J. W. (2017). "Gaussian process based nonlinear latent structure discovery in multivariate spike train data," in *Advances in Neural Information Processing Systems*, eds I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Curran Associates, Inc.), 3499–3508.

Wu, M. C., David, S. V., and Gallant, J. L. (2006). Complete functional characterization of sensory neurons by system identification. *Annu.*

Rev. Neurosci. 29, 477–505. doi: 10.1146/annurev.neuro.29.051605.113024

Yu, B. M., Cunningham, J. P., Santhanam, G., Ryu, S. I., Shenoy, K. V., and Sahani, M. (2009). "Gaussian-process factor analysis for low-dimensional single-trial analysis of neural population activity," in *Advances in Neural Information Processing Systems*, eds D. Koller, D. Schuurmans, Y. Bengio, and L. Bottou (Curran Associates, Inc.), 1881–1888.

Zhao, Y., and Park, I. M. (2017). Variational latent Gaussian process for recovering single-trial dynamics from population spike trains. *Neural Comput.* 29, 1293–1316. doi: 10.1162/NECO_a_00953

Zoltowski, D. M., and Pillow, J. W. (2018). "Scaling the Poisson GLM to massive neural datasets through polynomial approximations," in *Advances in Neural Information Processing Systems 31*, eds S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett (Curran Associates, Inc.), 3521–3531.

Check for updates

An Integrated Neuronal Model of Claustral Function in Timing the Synchrony Between Cortical Areas

Trichur R. Vidyasagar[1,2,3]* and Ekaterina Levichkina[1,4]

[1]Department of Optometry and Vision Science, University of Melbourne, Parkville, VIC, Australia, [2]Florey Institute of Neuroscience and Mental Health, Parkville, VIC, Australia, [3]Australian Research Council Centre of Excellence in Integrative Brain Function, University of Melbourne Node, Melbourne, VIC, Australia, [4]Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow, Russia

It has been suggested that the function of the claustrum (CL) may be to orchestrate and integrate the activity of the different cortical areas that are involved in a particular function by boosting the synchronized oscillations that occur between these areas. We propose here a model of how this may be done, thanks to the unique synaptic morphology of the CL and its excitatory and inhibitory connections with most cortical areas. Using serial visual search as an example, we describe how the functional anatomy of the claustral connections can potentially execute the sequential activation of the representations of objects that are being processed serially. We also propose that cross-frequency coupling (CFC) between low frequency signals from CL and higher frequency oscillations in the cortical areas will be an efficient means of CL modulating neural activity across multiple brain regions in synchrony. This model is applicable to the wide range of functions one performs, from simple object recognition to reading and writing, listening to or performing music, etc.

Keywords: claustrum, neural synchrony, cross-frequency coupling, visual cortex, visual search, attention

A REGION THAT INTEGRATES BRAIN ACTIVITY

For purposeful and useful interaction with the external world, the brain needs to integrate information processed in different parts of the nervous system, so that it can efficiently process sensory inputs, often from more than one modality, stored memories, emotional aspects of the situation, and executive and motor programmes needed for the chosen response. This requires the operation of many brain areas communicating with each other. Crick and Koch (2005) published a stimulating idea that in the claustrum (CL), the brain may have a central integrator essential for our unified sense of cognition and cohesive behavior. This insight was inspired by the anatomical connectivity between the CL and other brain regions and the synaptic organization within the nucleus itself. The CL connects reciprocally with almost every cortical area (Pearson et al., 1982; Tanné-Gariépy et al., 2002; Druga, 2014; Torgerson et al., 2015; Wang et al., 2017). Furthermore, CL has been found to be the most densely interconnected structure in the human brain (Torgerson et al., 2015), and its internal structure can facilitate rapid development of synchronized activity within adjacent activated regions of the CL (Crick and Koch, 2005; Smythies et al., 2014a; Kim et al., 2016). Crick and Koch (2005) suggested that the dendro-dendritic synapses in the CL, which could potentially include gap junctions (Shepherd and Greer, 1998), can rapidly transfer signals

arriving from different cortical areas. However, in the only study done in awake behaving macaques specifically aiming to record from multimodal neurones that would support an integrating function for single claustral cells, Remedios et al. (2010) found mainly unimodal sensory cells responding either to visual or auditory stimuli but not to both. Recent rodent studies of claustral circuitry have also shown only very weak connections between principal claustrocortical neurons (Kim et al., 2016). Smythies et al. (2012, 2014a,b) considering a few different hypotheses about how the CL may nevertheless be involved in integrating the activity across many parts of the brain, suggested that the most likely way the CL could exert its integrative function may not be by convergence of signals from various cortical areas on to single claustral cells, but rather by aiding cortical areas to amplify the synchrony between themselves. Saalmann et al. (2012) showed that the pulvinar does a comparable function in the maintenance of a working memory trace in a spatially cued object identification task. They demonstrated that the memory of the object location was maintained by a local cluster of pulvinar cells, as observed in the high degree of local spike-field coherence in the 8–15 Hz range and leading to almost zero-lag synchrony between visual areas V4 and TEO. This finding was supported by Zhou et al. (2016), who found a similar result prior to the appearance of the stimulus array in their paradigm. These synchronized oscillations were related to the maintenance of a memory trace that would be needed in the immediate future. Could the CL be doing something similar with regard to the actual processing and integration of sensory information and the behavioral response?

CLAUSTRUM COULD ENHANCE SYNCHRONIZED NEURAL OSCILLATIONS BETWEEN CORTICAL AREAS

A common principle of the mammalian brain that is being recognized as a fundamental mechanism driving its perception, cognition and behavior is the existence of periodic oscillations of neural activity amongst groups of active cells (e.g., Engel et al., 1991; Buzsáki et al., 2013; Buzsáki and Schomburg, 2015). Such rhythmic coordination in excitability is ubiquitous in the brain, but varying in its power, phase and frequency between brain regions and also between tasks. Almost every cortical activity involving processing of sensory information, memory, executive prerogatives and/or behavioral output inevitably engages multiple cortical areas communicating with each other and providing feedforward, feedback or modulatory signals. A plausible mechanism for such inter-areal communication is "communication through coherence" (Bastos et al., 2015), where rhythmic synchronization in one group of neurons leads to modulations in the input gain at synapses that they make on a second group of neurones. Such communication through coherence has been well documented by a number of studies through simultaneous recordings from two different cortical areas in awake macaques performing visual attention tasks (Buschman and Miller, 2007, 2009; Saalmann et al., 2007; Gregoriou et al., 2009).

Smythies et al. (2014a,b) suggested that when two cortical areas that are mutually connected and in a particular task begin to synchronize their activities, their connections to the CL first lead to rapid development of intraclaustral synchrony. These claustral regions are then believed to cause an increase in the synchrony between the two cortical regions through their efferents back to the respective cortical targets. While this proposition addresses the lack of multimodal neurones in the CL and yet ascribes to the CL a central integrative function, it opens up a number of new questions. Most importantly: (1) What is the relationship between the claustrocortical and cortico-cortical synchronies, in particular, do they occur at the same frequency? (2) What is the trigger for getting the CL involved and what terminates the synchrony generated in the cortex by the CL, without letting it evolve into a reverberating or even epileptiform discharge?

OUR HYPOTHESIS OF "PUNCTUATED NEURAL SYNCHRONY"

In this section, we outline a hypothesis for claustral function and illustrate it by applying it to "serial visual search". Visual search is not only a very common function our brains perform, but is also a widely studied task in both humans and non-human primates. In most variations of this paradigm, one searches for a target among a number of items in a visual scene, with which the target shares one or more features. Early visual search experiments by Treisman and Gelade (1980) led them to propose a "feature integration theory" to explain how we detect objects in a cluttered visual scene and also how we are able to bind the attributes of each object before identification. This highly influential model proposes that a "spotlight of attention" selects at a time one particular object in the visual scene to be processed in detail and then moves on to others until the target is found. As a neural correlate of the feature integration model, it has been proposed (Vidyasagar, 1999; Bullier, 2001) that the dorsal cortical stream and its top-down feedback to the primary visual cortex (area V1) and to ventral stream structures serially select, from a priority map in the posterior parietal cortex, one particular location for a short time (**Figure 1A**). This is then processed in detail by the ventral areas that deal with object recognition. Despite the functional localization in the primate brain with different areas and neurones being specialized for different attributes such as color and shape, the simultaneous processing of the attributes of only one object at any one time leads to the binding of features of that object alone. In doing this, serial search proceeds at a rate of 20–45 ms/item, depending upon task demands (Wolfe et al., 1998; Wolfe and Horowitz, 2004). This translates into largely a beta and low gamma frequency range (22.2–50 Hz). The main neurophysiological support for this claim arises from a number of studies: (1) Buschman and Miller (2009, 2010) show that covert shifts of attention in a visual search task is correlated with the cyclical oscillation of top-down prefrontal modulation of parietal activity occurring in the low gamma range; (2) there is a wealth of evidence for the role of lateral intraparietal area (LIP) in directing top-down attention to specific objects
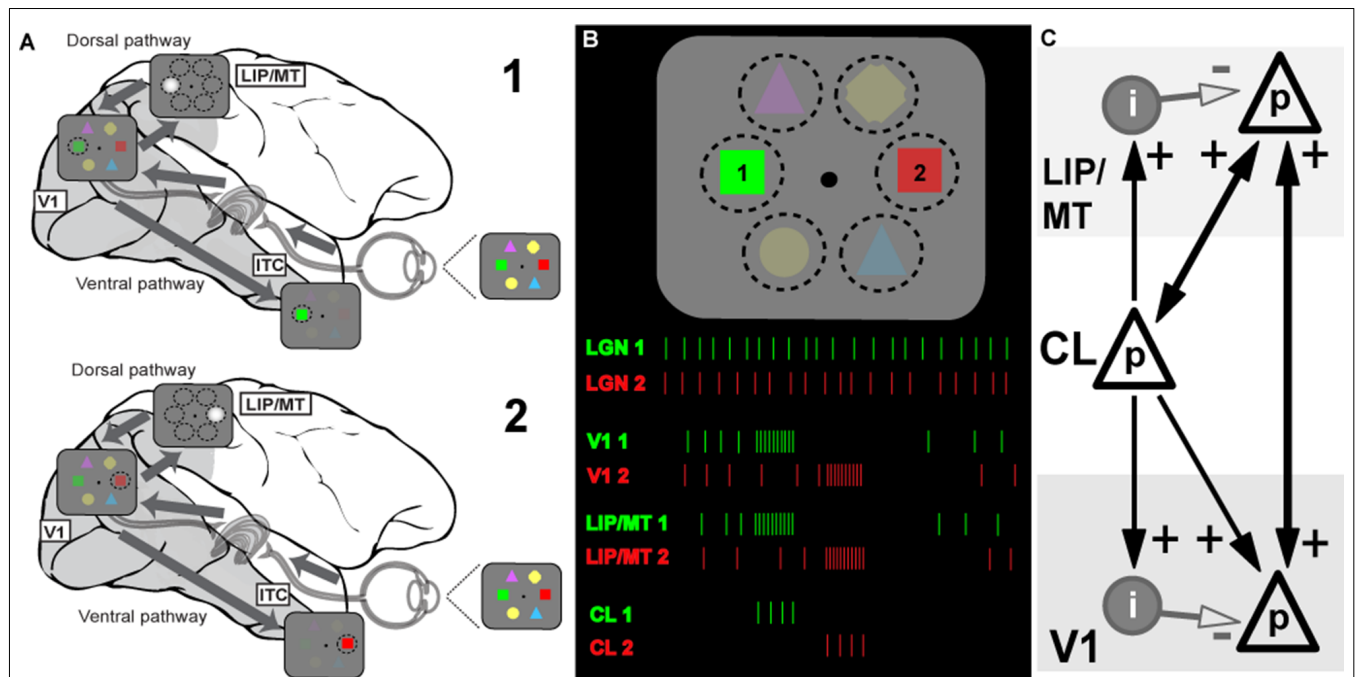
FIGURE 1 | Model of the information flow during visual search and the role of claustrum (CL) in orchestrating this process. **(A)** Schematic depiction of the signal processing occurring in dorsal and ventral visual streams during visual search. Visual stimulus array is shown at the right side of the panel (and in B). Due to differences in speed of transmission, visual information first reaches areas of the dorsal stream (lateral intraparietal area/middle temporal, LIP/MT) *via* the faster magnocellular pathway. Dorsal stream areas provide spatial feedback to primary visual cortex (V1) and ventral stream areas in the form of a spotlight of attention (represented by bright gray circle). This feedback arrives at V1 by the time the slower parvocellular-mediated information reaches it, and facilitates further processing of stimulus just for the part of the visual scene where attention is directed to. The process serves to limit information overload in ventral stream areas of the inferotemporal cortex (ITC) by processing one item at a time, and helps to solve the binding problem as well (for more details see Vidyasagar, 1999). During visual search, parts of visual space containing salient features are processed sequentially, as represented by stages 1 and 2 corresponding to attentive processing of green and red figures of the visual array, respectively. **(B)** Visual search task and putative neuronal activities in key brain areas: lateral geniculate nucleus (LGN), V1, dorsal stream (LIP) and CL. The same visual stimulus array is presented at the top, spike responses are shown below as the attentional spotlight is focussed first on target 1 (green) and then on target 2 (red). The initial volley of excitatory burst from CL neurons to LIP/MT and to V1 is followed by feedforward inhibition which terminates the processing of each stimulus. Sustained activity of LGN provides relatively constant input for processing, the dorsal stream organizes attentional spotlights, and CL determines timing of item-by-item processing during visual search. **(C)** Claustral connections with V1 and dorsal stream areas (LIP/MT). p refers to excitatory cell in layer 4 of the cortex and i represents an inhibitory interneuron. The strength of functional connections is shown by the thickness of the arrows.

(Bisley and Goldberg, 2003, 2010; Saalmann et al., 2007; Corbetta and Shulman, 2011; Meehan et al., 2017); (3) experiments in behaving macaques have shown that the top-down attentional feedback modulation of an early visual area, middle temporal (MT or area V5) by the parietal area, LIP is mediated by synchronized oscillations from LIP driving MT neurones at topographically corresponding locations, in the frequency range 25–45 Hz (Saalmann et al., 2007); and (4) though such cyclical modulation has not been directly demonstrated in the dorsal stream feedback to area V1, attentional and contextual modulation of V1 responses to visual inputs has long been amply demonstrated (Vidyasagar, 1998; Brefczynski and DeYoe, 1999; Ito and Gilbert, 1999; Gandhi et al., 1999; McAdams and Reid, 2005; Vidyasagar and Pigarev, 2007). Given the extensive neurophysiological evidence for synchronized neural oscillations in mediating interareal communication (Buschman and Miller, 2007; Saalmann et al., 2007, 2012; Gregoriou et al., 2009), it is not too speculative to suggest that the feedback to primary visual cortex is also likely to be mediated by such oscillations (Graboi and Lisman, 2003; Vidyasagar, 2013).

Extending the above argument, we propose that the CL's comprehensive reciprocal connections with almost all cortical areas and their unique internal morphology help to magnify the synchrony between cortical areas and also provide a behaviorally useful sequence of activation across the surface of the corresponding cortical areas, such as what is needed in tasks such as serial visual search. Claustral anatomy and its connectivity are likely to accomplish the above requirements. In **Figure 1C**, we show a simplified canonical circuitry which is applicable to any two or more cortical areas that are functionally connected to the CL in any particular situation, but here shown for a visual task. Taking serial visual search as example, we show on the right claustral efferents projecting to principal (p) cells in both V1 and the dorsal stream (here, marked as LIP/MT).

Afferents to input layers in cortical areas not only synapse on to the excitatory stellate and pyramidal cells, but also to local inhibitory interneurons. Studied most intensively in the primary visual cortex (Creutzfeldt and Ito, 1968; Ferster and Lindström, 1983; LeVay, 1986), such an input leads to a powerful and long-lasting inhibition. Such strong feed forward inhibition

(FFI) following on the heels of an excitatory input leads to aborting the excitatory response of the target cells after the initial volley (Bruno, 2011). While FFI has been shown to generate oscillations in a local network (Kremkow et al., 2010), FFI from one area to another, here from CL to V1, would serve another additional purpose, namely terminating the initial excitation.

Figure 1B shows how this may function in the case of serial visual search. In a typical visual search task, both engagement and disengagement from the items are essential and furthermore, they should occur sequentially, shifting from one item to another until the target is found, and all of this governed by task priorities. It is now believed (reviewed in Bisley and Goldberg, 2010) that area LIP has a continuously updated priority map that governs the allocation of top-down attentional signal. This priority map itself is updated from a number of inputs—especially task demands as dictated by prefrontal executive areas and saliency of the targets themselves (Ipata et al., 2009; Bisley and Goldberg, 2010). We propose that while the serial engagement of attention is determined simply by the pecking order in the priority map, the disengagement comes from the termination of the synchronized oscillations by the claustrocortical connections with areas that respond to the attributes of the object at the prioritized location. We suggest that such termination and thus the disengagement from the attended item is brought about by the inhibitory volley of the FFI circuit. Since such inhibition is long-lasting, it may also be the neural basis of "inhibition of return" (Wang and Klein, 2010), well-known in the visual search literature. Our proposed role of CL in facilitating top-down attentional modulation is consistent with results of recent experiments in rodents (Mathur, 2014; Goll et al., 2015; Atlan et al., 2018; White et al., 2018). Interestingly, CL not only receives selective top-down attentional influences from the cortex, for example from the anterior cingulate cortex (White et al., 2018), but it also plays a critical role in suppressing auditory distractors in a visual task (Atlan et al., 2018). Such a function is probably related to CL's role in helping to distinguish between relevant and irrelevant items as in a typical search task.

For attentive serial search to work in the fashion described above, we expect that any reciprocal connection from V1 to CL is weak or non-existent. As described earlier, serial search requires moving the spotlight of attention from one item of the scene to another until the target is found. Object recognition is known to occur largely in the ventral stream and it is believed to be facilitated by top-down modulation of incoming visual signals by feedback from the dorsal stream (Vidyasagar, 1999; Bullier, 2001). Once visual attention gets focussed on one object by the spotlight of attention, the CL may play little role in the more detailed processing by the ventral cortical areas. Finding the target would abort the FFI from the CL and the activity in V1 and the corresponding topographic locations in the various cortical areas would continue under focussed attention. Furthermore, if activity related to object locations are supposed to be "serially highlighted" for further processing by extra-striate areas such as LIP, MT, V4 and TEO for ultimate binding of the attributes of the object, such a scheme would be defeated if there are strong signals from every item to the CL, triggering reciprocal synchronizing

volleys. In fact, many studies on the CL, while describing the widespread afferent connections from the CL to most association areas and the prefrontal cortex have emphasized the uncertainty of the projection from the primary sensory areas, including V1 in the primate (Druga, 2014; reviewed in Smythies et al., 2014a). There is also a cautionary note about the effectiveness of the V1 (area 17) to CL projection that has been described in the cat. LeVay and Sherk (1981), who studied connections between visual areas and CL in the cat, found that area 17 cells projecting to CL were just 3.5% of layer 6 cells and these were found predominantly in the peripheral rather than central visual field representation, whereas the claustral projections to area 17 were much heavier. Sherman and Guillery (2011) state that the layer 6 cells that project subcortically are class 2 glutamatergic cells that do not produce much spiking activity but only modulate responses mainly through metabotropic postsynaptic receptors. Thus, the claustral synchrony may get initiated and sustained, not so much by the sensory input to primary sensory areas, but rather by activity in higher areas such as LIP. Thereafter, as the enhanced synchrony between the representations of a particular object in different cortical areas (in **Figure 2**, V1 and LIP/MT) develops and then dies down with its termination by FFI, the next most salient location in LIP synchronizes with V1 and the corresponding locations in the CL also get activated and a new cycle of enhanced synchrony starts, to be in its turn terminated by the subsequent FFI.

Recent studies of the rodent CL have demonstrated the strong inhibitory influence that optogenetic stimulation of claustral outputs could have on cortical areas, namely on unit responses in the anterior cingulate cortex (White et al., 2018), the prefrontal cortex (Jackson et al., 2018; Narikiyo et al., 2018) and the auditory cortex (Atlan et al., 2018). We suggest that these inhibitory volleys represent the FFI needed to terminate activity in target areas as described above in our scheme. It is noteworthy that in all of these studies, the optogenetic excitation was
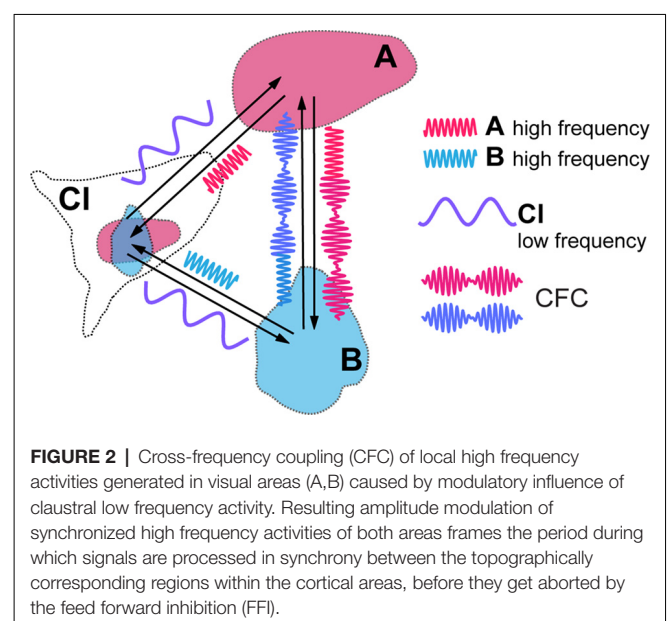

FIGURE 2 | Cross-frequency coupling (CFC) of local high frequency activities generated in visual areas (A,B) caused by modulatory influence of claustral low frequency activity. Resulting amplitude modulation of synchronized high frequency activities of both areas frames the period during which signals are processed in synchrony between the topographically corresponding regions within the cortical areas, before they get aborted by the feed forward inhibition (FFI).

effective in causing the inhibition in target cortical areas, but the initial excitatory response in the target area was rather weak (Narikiyo et al., 2018). This may be attributed to two factors: (i) the optogenetic stimulation does not resemble the usual synchronized oscillatory activity that may be needed to cause the excitatory oscillations as described above and in the section below on cross-frequency coupling (CFC) and (ii) the excitatory response would require temporal simultaneity of oscillatory inputs from other cortical areas.

Recent rodent studies have also elucidated a claustral circuitry that could be ideally suited to our proposed function of the CL (Kim et al., 2016, see their Figure 8), by possibly enabling another FFI circuit within the CL itself. While corticoclaustral inputs target individual claustrocortical (ClaC) cells monosynaptically and there are few direct connections among these principal, claustrocortical cells, the cortical inputs to CL provide strong stimulation to the parvalbumin (PV)-positive inhibitory interneurons, which are themselves strongly interconnected *via* both electrical and chemical synapses (Kim et al., 2016). This leads to a situation where synchronous activation signals from two different cortical areas to their reciprocally connected ClaC cells would set off a neural synchrony between the cortical areas and the CL, soon to be followed by an inhibitory volley that suppresses the claustral outputs.

Finally, when the target in a visual search task is found, the termination of all activity in the CL and the search itself may be brought about by stimulation of the kappa opioid receptors, the mRNA for which is particularly plentiful in the CL (Mansour et al., 1994). The high density of these receptors on claustral cells is a striking finding that needs particular consideration in any model of claustral function. The possible role of this receptor system in the larger integrative functions has been pointed out (Stiefel et al., 2014), since such receptor stimulation inhibits the release of GABA (Hjelmstad and Fields, 2003; Li et al., 2012) which in turn would disrupt the generation of oscillations within the CL and the claustral amplification of the synchrony between cortical areas. Activation of the kappa receptors inhibits both glutamate and GABA transmission (Hjelmstad and Fields, 2003), thus practically stopping excitatory activity as well as disrupting oscillations. We believe that a match between an object brought under the roving spotlight of attention and the representation of the expected object may abort the visual search through its effect on claustral kappa opioid receptors. While the kappa opioid system may be generally known for its dysphoric effects, particularly in producing the aversive and depressive effects in the case of drug abuse (Lalanne et al., 2014), the evolutionary reason for the kappa receptors are not likely to be related to drug addiction. Natural opioids acting on mu opioid receptor (MOR) and kappa receptors are known to lead to opposing effects in rats performing a behavioral task, the former to reinforcement of the related behavior and the latter to its termination (Shippenberg and Herz, 1986). Though stimulation of kappa receptors in the ventral tegmental area may be related to motivational and hedonic aspects (Spanagel et al., 1992), similar stimulation in other areas may have effects depending upon the function of those respective areas.

Thus, their primary role may be simply in aborting neural synchrony in local circuits through their action on GABEergic transmission, besides the inhibition of the excitatory activity itself. We propose that until the visual search is completed, there is little stimulation of the claustral kappa-opioid receptors, but a specific input to the CL on finding the target, possibly from the prefrontal regions which are heavily linked to the CL (Reser et al., 2014), may disrupt neural oscillations in the CL and consequently its amplification of synchrony in various cortical regions.

Our model of claustral control of visual search is one convenient example for what we believe to be a description of claustral function in general. We believe that the proposed role of CL in sequencing neuronal activity is not restricted to the visual modality, but in line with its widespread cortical connections, CL can potentially modulate activity in all sensory cortices, association areas and also motor areas. Thus, we hypothesize that the CL might be instrumental in not only in binding the activity of different cortical regions by enhancing their synchrony, but also organizing all cortex-mediated processes in a sequential manner, as for example in language comprehension, language production and in organizing complex motor programs.

CLAUSTRAL MODULATION OF OTHER BRAIN AREAS THROUGH CROSS-FREQUENCY COUPLING

CFC is being recognized as an efficient means of communication between two cortical areas and it is likely to play a critical role in mediating working memory and in enabling learning (Canolty and Knight, 2010; Lisman and Jensen, 2013; Hyafil et al., 2015). Blood-oxygen-level dependent (BOLD) connectivity between areas is best predicted by low frequency oscillations that determine the amplitude of gamma frequencies (Wang et al., 2012). Thus, in the above example, in target cortical areas such as LIP, MT and V1, the amplitudes of a higher, such as high beta or gamma, frequency rhythm may be modulated by, and thus nested within, a lower frequency, for instance theta, alpha or low beta, at which claustral efferents send out their modulating signals to their targets (**Figure 2**). We expect that each cycle of the low frequency signal from CL would allow sufficient number of high frequency cycles at its target areas to synchronize before the excitatory volley gets aborted by the FFI. While electrical stimulation of lateral geniculate nucleus (LGN) leads to disynaptically mediated inhibitory post-synaptic potentials in stellate cells in layer 4 of the primary visual cortex within a few milliseconds (Creutzfeldt and Ito, 1968; Ferster and Lindström, 1983), with visual stimulation the inhibition seen in intracellular recordings from the cat striate cortex develops gradually over many tens of milliseconds (Pei et al., 1994; Volgushev et al., 1995: Ringach et al., 1997). Both with such visual stimulation and with electrical stimulation (Viswanathan et al., 2011), the inhibition can however last many hundreds of milliseconds. Strong FFI caused by CL stimulation and mediated *in vivo* by relatively slow neuropeptide Y interneurons was also described in the prefrontal areas of rodents by Jackson et al. (2018), with the

excitation/inhibition ratio of cortical pyramidal cells equalling just 0.25. Though one is yet to see similar studies done in the case of the primate CL, the window of opportunity for neural synchrony between relevant cortical regions to be amplified by claustral output is likely to be defined by the time course of the FFI circuit. It is possible that this time course may also be modulated by task demands and the state of vigilance.

The cyclical facilitation of processing of incoming visual signals in V1 would mean that sensitivity to visual stimuli could show periodic fluctuation, as indeed they do (Busch et al., 2009; Mathewson et al., 2009; VanRullen and Dubois, 2011). CFC with nested frequencies may also be critical for processing of stimuli at multiple temporal rates, such as graphemes/phonemes, and syllables and words during reading and speech perception (Graboi and Lisman, 2003; Vidyasagar, 2013). Through CFC, claustral output at one low frequency (delta, theta, alpha, or low beta) can modulate a range of oscillation frequencies (high beta or gamma) at cortical areas that are connected to each other in a task such as reading or visual search. **Figure 2** is a simplified diagram of how this might function in the case of CL boosting synchrony between LIP/MT and V1. At this stage, it is too premature to speculate at what frequency the claustral assembly oscillates. It may be either always at the same frequency which is determined by its own morphology and resonance frequency or dictated by the area that triggers the synchrony in the first place or even under an executive command from the prefrontal cortex.

OUTSTANDING QUESTIONS FOR FUTURE STUDIES

The model leads to a number of testable predictions. The following are some of the main questions for study.

1. In tasks such as visual search, the model predicts neural activity in CL driving synchronized activities in relevant cortical regions.
2. Claustral influence on cortical regions would exhibit two stages: an initial excitatory oscillation followed by strong inhibition.
3. The time course of the FFI from CL on cortical areas needs to be ascertained to test whether it permits synchrony between cortical areas.
4. Is the low frequency volley from CL mediating CFC fixed or is it dynamically modified by task demands?
5. In a visual search task, is there a roving wave of synchrony across the CL as the animal performs a search task, as the corresponding topographic locations in the CL serially facilitate the scan of the spotlight of attention?

6. Is there a rapid termination of intraclaustral synchrony and stimulation of GABAergic neurons as soon as the target is found?

Some of these questions need to be addressed in awake non-human primates. So far, with rare exceptions (Remedios et al., 2010, 2014) the primate CL has defied functional studies, due to its shape and anatomical location, but it is possible that with newer emerging techniques, the experiments are feasible.

CONCLUSION

Our hypothesis suggests the existence of a functional circuit by which CL could play a vital role in communication between cortical areas by enhancing both the synchrony between cortical areas as well the amplitude of oscillations. The scheme has the advantage that though the connections between cortical areas themselves may not be structurally and functionally strong to develop enough synchrony, the boost given by the CL can help them to attain a degree of synchrony that will be functionally useful. Critical to this function is the unique claustral morphology (Kim et al., 2016) and the FFI circuit both within the CL and in its cortical targets, which are features considered to be characteristic of a system designed to amplify correlated neuronal activity (Bruno, 2011; Hu et al., 2014). The metaphor that Crick and Koch (2005) thought of, that the CL is like the conductor of an orchestra, is apt in more ways than one. In short, the punctuated synchrony we propose is akin to the conductor of an orchestra co-ordinating and inspiring a harmonious and smoothly punctuated symphony. In short, it is a *conductor of the synchrony* between cortical areas.

AUTHOR CONTRIBUTIONS

TV was responsible for the basic idea proposed in the article and drafting the first version. He also conducted some of the experiments that underpin crucial elements of the proposed theory. EL assisted in developing the basic idea to fit into a diagrammatic scheme and provided critical input towards the article's intellectual content. EL also enhanced comprehension of the article with revisions and the illustrations.

FUNDING

REFERENCES

Atlan, G., Terem, A., Peretz-Rivlin, N., Sehrawat, K., Gonzales, B. J., Pozner, G., et al. (2018). The claustrum supports resilience to distraction. *Curr. Biol.* 28, 2752.e7–2762.e7. doi: 10.1016/j.cub.2018.06.068

Bastos, A. M., Vezoli, J., and Fries, P. (2015). Communication through coherence with inter-areal delays. *Curr. Opin. Neuribiol.* 31, 173–180. doi: 10.1016/j.conb.2014.11.001

Bisley, J. W., and Goldberg, M. E. (2003). Neuronal activity in the lateral intraparietal area and spatial attention. *Science* 299, 81–86. doi: 10.1126/science.1077395

Bisley, J. W., and Goldberg, M. E. (2010). Attention, intention and priority in the parietal lobe. *Annu. Rev. Neurosci.* 33, 1–21. doi: 10.1146/annurev-neuro-060909-152823

Brefczynski, J. A., and DeYoe, E. A. (1999). A physiological correlate of the "spotlight" of visual attention. *Nat. Neurosci.* 2, 370–374. doi: 10.1038/7280

Bruno, R. M. (2011). Synchrony in sensation. *Curr. Opin. Neurobiol.* 21, 701–708. doi: 10.1016/j.conb.2011.06.003

Bullier, J. (2001). Integrated model of visual processing. *Brain Res. Brain Res. Rev.* 36, 96–107. doi: 10.1016/s0165-0173(01)00085-6

Busch, N. A., Dubois, J., and VanRullen, R. (2009). The phase of ongoing EEG oscillations predicts visual perception. *J. Neurosci.* 29, 7869–7876. doi: 10.1523/JNEUROSCI.0113-09.2009

Buschman, T. J., and Miller, E. K. (2007). Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. *Science* 315, 1860–1862. doi: 10.1126/science.1138071

Buschman, T. J., and Miller, E. K. (2009). Serial, covert shifts of attention in visual search are reflected by the frontal eye fields and correlated with population oscillations. *Neuron* 63, 386–396. doi: 10.1016/j.neuron.2009.06.020

Buschman, T. J., and Miller, E. K. (2010). Shifting the spotlight of attention: evidence for discrete computations in cognition. *Front. Hum. Neurosci.* 4:194. doi: 10.3389/fnhum.2010.00194

Buzsáki, G., and Schomburg, E. V. (2015). What does gamma coherence tell us about inter-regional neural communication? *Nat. Neurosci.* 18, 484–489. doi: 10.1038/nn.3952

Buzsáki, G., Logothetis, N., and Singer, W. (2013). Scaling brain size, keeping timing: evolutionary preservation of brain rhythms. *Neuron* 80, 751–764. doi: 10.1016/j.neuron.2013.10.002

Canolty, R. T., and Knight, R. T. (2010). The functional role of cross-frequency coupling. *Trends Cogn. Sci.* 14, 506–515. doi: 10.1016/j.tics.2010.09.001

Corbetta, M., and Shulman, G. L. (2011). Spatial neglect and attention networks. *Annu. Rev. Neurosci.* 34, 569–599. doi: 10.1146/annurev-neuro-061010-113731

Creutzfeldt, O. D., and Ito, M. (1968). Functional synaptic organization of primary visual cortex neurones in the cat. *Exp. Brain Res.* 6, 324–352. doi: 10.1007/bf00233183

Crick, F. C., and Koch, C. (2005). What is the function of the claustrum? *Phil. Trans. Roy. Soc. Lond. B Biol. Sci.* 36, 1271–1279. doi: 10.1098/rstb.2005.1661

Druga, R. (2014). "The structure and connections of the claustrum," in *The Claustrum*, ed. J. Smythies (San Diego: Elsevier), 29–84.

Engel, A. K., König, P., Kreiter, A. K., and Singer, W. (1991). Interhemispheric synchronization of oscillatory neuronal responses in cat visual cortex. *Science* 88, 1177–1179. doi: 10.1126/science.252.5009.1177

Ferster, D., and Lindström, S. (1983). An intracellular analysis of geniculo-cortical connectivity in area 17 of the cat. *J. Physiol.* 342, 181–215. doi: 10.1113/jphysiol.1983.sp014846

Gandhi, S. P., Heeger, D. J., and Boynton, G. M. (1999). Spatial attention affects brain activity in human primary visual cortex. *Proc. Nat. Acad. Sci. U S A* 96, 3314–3319. doi: 10.1073/pnas.96.6.3314

Goll, Y., Atlan, G., and Citri, A. (2015). Attention: the claustrum. *Trends Neurosci.* 38, 486–495. doi: 10.1016/j.tins.2015.05.006

Graboi, D., and Lisman, J. (2003). Recognition by top-down and bottom-up processing in cortex: the control of selective attention. *J. Neurophysiol.* 90, 798–810. doi: 10.1152/jn.00777.2002

Gregoriou, G. G., Gotts, S. J., Zhou, H., and Desimone, R. (2009). Long-range neural coupling through synchronization with attention. *Science* 176, 35–45. doi: 10.1016/s0079-6123(09)17603-3

Hjelmstad, G. O., and Fields, H. L. (2003). Kappa opioid receptor activation in the nucleus accumbens inhibits glutamate and GABA release through different mechanisms. *J. Neurophysiol.* 89, 2389–2395. doi: 10.1152/jn.01115.2002

Hu, H., Gab, J., and Jonas, P. (2014). Interneurons, fast-spiking, parvalbumin+ GABAergic interneurons: from cellular design to microcircuit function. *Science* 345:1255263. doi: 10.1126/science.1255263

Hyafil, A., Giraud, A. L., Fontolan, L., and Gutkin, B. (2015). Neural cross-frequency coupling: connecting architectures, mechanisms and functions. *Trends Neurosci.* 38, 725–740. doi: 10.1016/j.tins.2015.09.001

Ipata, A. E., Gee, A. L., Bisley, J. W., and Goldberg, M. E. (2009). Neurons in the lateral intraparietal area create a priority map by the combination of disparate signals. *Exp. Brain Res.* 192, 479–488. doi: 10.1007/s00221-008-1557-8

Ito, M., and Gilbert, C. D. (1999). Attention modulates contextual influences in the primary visual cortex of alert monkeys. *Neuron* 22, 593–604. doi: 10.1016/s0896-6273(00)80713-8

Jackson, J., Karnani, M. M., Zemelman, B. V., Burdakov, D., and Lee, A. K. (2018). Inhibitory control of prefrontal cortex by the claustrum. *Neuron.* 99, 1029.e4–1039.e4. doi: 10.1016/j.neuron.2018.07.031

Kim, J., Matney, C. J., Roth, R. H., and Brown, S. P. (2016). Synaptic organization of the neuronal circuits of the claustrum. *J. Neurosci.* 36, 773–784. doi: 10.1523/jneurosci.3643-15.2016

Kremkow, J., Perrinet, L. U., Masson, G. S., and Aertsen, A. (2010). Functional consequences of correlated excitatory and inhibitory conductances in cortical networks. *J. Comput. Neurosci.* 28, 579–594. doi: 10.1007/s10827-010-0240-9

Lalanne, L., Ayranci, G., Kieffer, B. L., and Lutz, P. E. (2014). The kappa opioid receptor: from addiction to depression and back. *Front. Psychiatry* 5:170. doi: 10.3389/fpsyt.2014.00170

LeVay, S. (1986). Synaptic organization of claustral and geniculate afferents to the visual cortex of the cat. *J. Neurosci.* 6, 3564–3575. doi: 10.1523/jneurosci.06-12-03564.1986

LeVay, S., and Sherk, H. (1981). The visual claustrum of the cat. I. Structure and connections. *J. Neurosci.* 1, 956–980. doi: 10.1126/science.7209525

Li, C., Pleil, K. E., Stamatakis, A. M., Busan, S., Vong, L., Lowell, B. B., et al. (2012). Presynaptic inhibition of gamma-aminobutyric acid release in the bed nucleus of stria terminalis by kappa opioid receptor signalling. *Biol. Pysch.* 71, 725–732. doi: 10.1016/j.biopsych.2011.11.015

Lisman, J. E., and Jensen, O. (2013). The Theta-Gamma neural code. *Neuron* 77, 1002–1016. doi: 10.1016/j.neuron.2013.03.007

Mansour, A., Fox, C. A., Burke, S., Meng, F., Thompson, R. C., Akil, H., et al. (1994). Mu, delta and kappa opioid receptor mRNA expression in the rat CNS: an *in situ* hybridization study. *J. Comp. Neurol.* 350, 412–438. doi: 10.1002/cne.903500307

Mathewson, K. E., Gratton, G., Fabiani, M., Beck, D. M., and Ro, T. (2009). To see or not to see: prestimulus alpha phase predicts visual awareness. *J. Neurosci.* 29, 2725–2732. doi: 10.1523/jneurosci.3963-08.2009

Mathur, B. N. (2014). The claustrum in review. *Front. Syst. Neurosci.* 8:48. doi: 10.3389/fnsys.2014.00048

McAdams, C. J., and Reid, R. C. (2005). Attention modulates the responses of simple cells in monkey primary visual cortex. *J. Neurosci.* 25, 11023–11033. doi: 10.1523/JNEUROSCI.2904-05.2005

Meehan, T. P., Bressler, S. L., Tang, W., Astafiev, S. V., Sylvester, C. M., Shulman, G. L., et al. (2017). Top-down cortical interactions in visuospatial attention. *Brain Struct. Funct.* 222, 3127–3145. doi: 10.1007/s00429-017-1390-6

Narikiyo, K., Mizuguchi, R., Ajima, A., Mitsui, S., Shiozaki, M., Hamanaka, H., et al. (2018). The claustrum coordinates cortical slow-wave activity. *BioRxiv* [preprint]. doi: 10.1101/286773

Pearson, R. C. A., Brodal, P., Gatter, K. C., and Powell, T. P. (1982). The organization of the connections between the cortex and the claustrum in the monkey. *Brain Res.* 234, 435–441. doi: 10.1016/0006-8993(82)90883-6

Pei, X., Vidyasagar, T. R., Volgushev, M., and Creutzfeldt, O. D. (1994). Receptive field analysis and orientation selectivity of postsynaptic potentials of simple cells in cat visual cortex. *J. Neurosci.* 14, 7130–7140. doi: 10.1523/jneurosci.14-11-07130.1994

Remedios, R., Logothetis, N. K., and Kayser, C. (2010). Unimodal responses prevail in multisensory claustrum. *J. Neurosci.* 30, 12902–12907. doi: 10.1523/jneurosci.2937-10.2010

Remedios, R., Logothetis, N. K., and Kayser, C. (2014). A role of the claustrum in auditory scene analysis by reflecting sensory change. *Front. Syst. Neurosci.* 8:44. doi: 10.3389/fnsys.2014.00044

Reser, D. H., Richardson, K. E., Montibeller, M. O., Zhao, S., Chan, J. M., Soares, J. G., et al. (2014). Claustrum projections to prefrontal cortex in the capuchin monkey (Cebus apella). *Front. Syst. Neurosci.* 8:123. doi: 10.3389/fnsys.2014.00123

Ringach, D. L., Hawken, M. J., and Dhapley, R. M. (1997). Dynamics of orientation tuning in macaque primary visual cortex. *Nature* 387, 281–284. doi: 10.1038/387281a0

Saalmann, Y. B., Pigarev, I. N., and Vidyasagar, T. R. (2007). Neural mechanisms of visual attention: how top-down feedback highlights relevant locations. *Science* 316, 1612–1615. doi: 10.1126/science.1139140

Saalmann, Y. B., Pinsk, M. A., Wang, L., Li, X., and Kastner, S. (2012). The pulvinar regulates information transmission between cortical areas based on attention demands. *Science* 337, 753–756. doi: 10.1126/science.1223082

Shepherd, G. M., and Greer, C. A. (1998). "Olfactory bulb," in *The Synaptic Organization of the Brain*, ed. G. M. Shepherd (New York, NY, USA: Oxford University Press), 159–204.

Sherman, S. M., and Guillery, R. W. (2011). Distinct functions for direct and transthalamic corticocortical connections. *J. Neurophysiol.* 106, 1068–1077. doi: 10.1152/jn.00429.2011

Shippenberg, T. S., and Herz, A. (1986). Differential effects of mu and kappa opioid systems on motivational processes. *NIDA Res. Monogr.* 75, 563–566. doi: 10.1037/e471872004-001

Smythies, J., Edelstein, L., and Ramachandran, V. (2012). Hypotheses relating to the function of the claustrum. *Front. Integr. Neurosci.* 6:53. doi: 10.3389/fnint.2012.00053

Smythies, J., Edelstein, L., and Ramachandran, V. (2014a). "Hypothesis relating to the function of the claustrum," in *The Claustrum*, ed. J. Smythies (San Diego: Elsevier), 299–352.

Smythies, J., Edelstein, L., and Ramachandran, V. (2014b). Hypothesis relating to the function of the claustrum II: does the claustrum use frequency codes? *Front Integr Neurosci.* 8:7. doi: 10.3389/fnint.2014.00007

Spanagel, R., Herz, A., and Shippenberg, T. S. (1992). Opposing tonically active endogenous opioid systems modulate the mesolimbic dopaminergic pathway. *Proc. Natl. Acad. Sci. U S A* 89, 2046–2050. doi: 10.1073/pnas.89.6.2046

Stiefel, K. M., Merrifield, A., and Holcombe, A. O. (2014). Claustrum's proposed role in consciousness is supported by the effect and target localization of saliva divinorum. *Front. Integr. Neurosc.* 26, 8–20. doi: 10.3389/fnint.2014.00020

Tanné-Gariépy, J., Boussaoud, D., and Rouiller, E. M. (2002). Projections of the claustrum to the primary motor, premotor and prefrontal cortices in the macaque monkey. *J. Comp. Neurol.* 454, 140–157. doi: 10.1002/cne.10425

Torgerson, C. M., Irimia, A., Goh, S. Y., and Van Horn, J. D. (2015). The DTI connectivity of the human claustrum. *Hum. Brain Mapp.* 36, 827–838. doi: 10.1002/hbm.22667

Treisman, A. M., and Gelade, G. (1980). A feature-integration theory of attention. *Cogn. Psych.* 12, 97–136. doi: 10.1016/0010-0285(80)90005-5

VanRullen, R., and Dubois, J. (2011). The psychophysics of brain rhythms. *Front. Psychol.* 2:203. doi: 10.3389/fpsyg.2011.00203

Vidyasagar, T. R. (1998). Gating of neuronal responses in macaque primary visual cortex by an attentional spotlight. *Neuroreport* 9, 1947–1952. doi: 10.1097/00001756-199806220-00006

Vidyasagar, T. R. (1999). A neuronal model of attentional spotlight: parietal guiding the temporal. *Brain Res. Rev.* 30, 66–76. doi: 10.1016/s0165-0173(99)00005-3

Vidyasagar, T. R. (2013). Reading into neuronal oscillations in the visual system: implications for developmental dyslexia. *Front. Hum. Neurosci.* 7:811. doi: 10.3389/fnhum.2013.00811

Vidyasagar, T. R., and Pigarev, I. N. (2007). Gating of neuronal responses in macaque primary visual cortex in a memory task. *Eur. J. Neurosci.* 25, 2547–2557.

Viswanathan, S., Jayakumar, J., and Vidyasagar, T. R. (2011). Role of feedforward geniculate inputs in the generation of orientation selectivity in the cat's primary visual cortex. *J. Physiol.* 589, 2349–2361. doi: 10.1113/jphysiol.2010.202317

Volgushev, M., Vidyasagar, T. R., and Pei, X. (1995). Dynamics of the orientation tuning of postsynaptic potentials in the cat visual cortex. *Vis. Neurosci.* 12, 621–628. doi: 10.1017/s0952523800008919

Wang, L., Saalmann, Y. B., Pinsk, M. A., Arcaro, M. J., and Kastner, S. (2012). Electrophysiological low-frequency coherence and cross-frequency coupling contribute to BOLD connectivity. *Neuron* 76, 1010–1020. doi: 10.1016/j.neuron.2012.09.033

Wang, Q., Ng, L., Harris, J. A., Feng, D., Li, Y., Royall, J. J., et al. (2017). Organization of the connections between claustrum and cortex in the mouse. *J. Comp. Neurol.* 525, 1317–1346. doi: 10.1002/cne.24047

Wang, Z., and Klein, R. M. (2010). Searching for inhibition of return in visual search. *Vis. Res.* 50, 220–228. doi: 10.1016/j.visres.2009.11.013

White, M. G., Panicker, M., Mu, C., Carter, A. M., Roberts, B. M., Dharmasri, P. A., et al. (2018). Anterior cingulate cortex input to the claustrum is required for top-down action control. *Cell Rep.* 22, 84–95. doi: 10.1016/j.celrep.2017.12.023

Wolfe, J. M., and Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nat. Rev. Neurosci.* 5, 495–501. doi: 10.1038/nrn1411

Wolfe, J. M., O'Neill, P., and Bennett, S. C. (1998). Why are there eccentricity effects in visual search? Visual and attentional hypotheses. *Percept. Psychophys.* 60, 140–156. doi: 10.3758/bf03211924

Zhou, H., Schafer, R. J., and Desimone, R. (2016). Pulvinar-Cortex Interactions in Vision and Attention. *Neuron* 89, 209–220. doi: 10.1016/j.neuron.2015.11.034

Mixed Spatial and Movement Representations in the Primate Posterior Parietal Cortex

Kostas Hadjidimitrakis [1,2†], Sophia Bakola [1,2*†], Yan T. Wong [1,3*†] and Maureen A. Hagan [1,2†]

[1]Department of Physiology, Monash University, Clayton, VIC, Australia, [2]Australian Research Council Centre of Excellence for Integrative Brain Function, Monash University Node, Clayton, VIC, Australia, [3]Department of Electrical and Computer Science Engineering, Monash University, Clayton, VIC, Australia

The posterior parietal cortex (PPC) of humans and non-human primates plays a key role in the sensory and motor transformations required to guide motor actions to objects of interest in the environment. Despite decades of research, the anatomical and functional organization of this region is still a matter of contention. It is generally accepted that specialized parietal subregions and their functional counterparts in the frontal cortex participate in distinct segregated networks related to eye, arm and hand movements. However, experimental evidence obtained primarily from single neuron recording studies in non-human primates has demonstrated a rich mixing of signals processed by parietal neurons, calling into question ideas for a strict functional specialization. Here, we present a brief account of this line of research together with the basic trends in the anatomical connectivity patterns of the parietal subregions. We review, the evidence related to the functional communication between subregions of the PPC and describe progress towards using parietal neuron activity in neuroprosthetic applications. Recent literature suggests a role for the PPC not as a constellation of specialized functional subdomains, but as a dynamic network of sensorimotor loci that combine multiple signals and work in concert to guide motor behavior.

Keywords: eye movements, reaching, grasping, PPC, posterior parietal cortex, movement planning

INTRODUCTION

Humans and non-human primates make skillful reaching-to-grasping movements that are tightly coordinated in space and time (Jeannerod et al., 1995). Moreover, eye movements often accompany every day actions towards objects, supplying information about object identity and location, and guiding arm movements (Johansson et al., 2001; Land and Hayhoe, 2001; Hayhoe et al., 2003). Contemporary research has established that the posterior parietal cortex (PPC) is involved in the representation of spatial information and goal-directed behavior using different motor effectors (Husain and Nachev, 2007; Andersen and Cui, 2009). Since the original unified view of PPC as a "command apparatus for the operation of the limbs, hands and eyes" (Mountcastle et al., 1975), anatomical, neurophysiological and neuroimaging evidence has ascribed the neural encoding of looking, reaching and grasping actions to distinct PPC sectors (Rizzolatti and Matelli, 2003; Vesia and Crawford, 2012; Andersen et al., 2014).
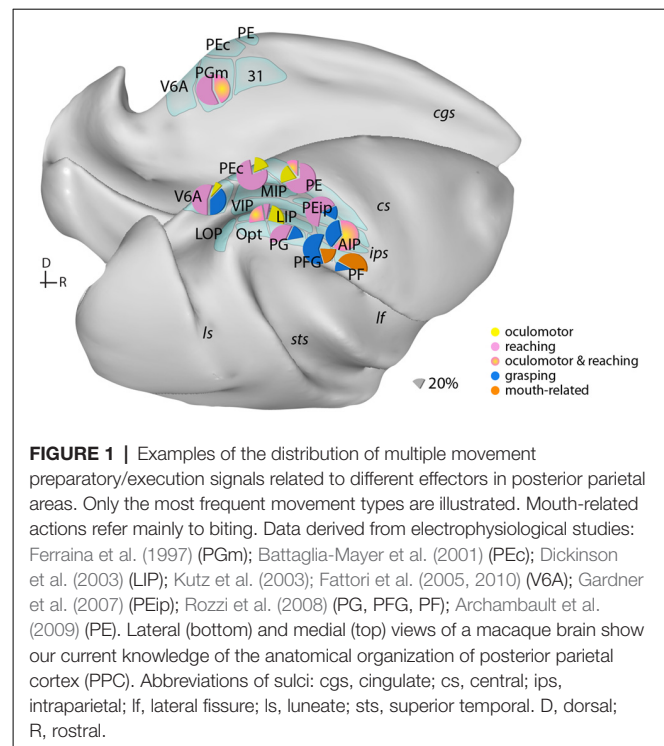
At the same time, numerous studies have shown convergence of eye-, arm- and/or hand-related signals, both within single PPC sectors and at the level of individual cells, although which of these signals play a casual role in defining functional specificity would require future investigations. Recent research findings raise several issues regarding the potential substrates of distinct

movements in parietal cortex and the information flow between the various PPC sectors. Here, we outline evidence, mainly from non-human primate anatomical and neurophysiological studies, for the rich variety of signals carried by PPC neurons related to movement guidance that suggests a more widespread representation of movement variables than previously assumed. From a clinical perspective, the diverse representation of signals from parietal cortex may prove useful for the design of more efficient neuroprosthetic devices for patients who cannot reach and grasp objects either because of loss of arms or lesions of the motor pathways.

ANATOMICAL ORGANIZATION OF THE POSTERIOR PARIETAL CORTEX

The PPC is composed of several areas that vary in histological features and connections with other parts of the brain. Definitions of areas have evolved over time from the historical assignment of posterior parietal fields to areas 5 and 7 of Brodmann to more refined schemes (e.g., **Figure 1**) but, despite general consensus on the number and characteristics of individual areas, maps produced by different groups vary widely and functional subdivisions do not always appear to respect architectonic boundaries (e.g., Savaki et al., 2010; Arcaro et al., 2011; Seelke et al., 2012). Nonetheless, in non-human primates, the anatomical organization of PPC is shaped by the relative influence of sensorimotor input to different areas. Segregated projections from the motor control centers in the frontal lobe are distributed along the dorsal-ventral extent of PPC. Primary motor cortex connects mainly to the parietal convexity (PE) and rostral parts of the medial bank of the intraparietal sulcus (IPS; PEip). Caudal superior and medial parietal areas (V6A, MIP, PEc, 31) connect preferentially with parts of dorsal premotor cortex, whereas inferior parietal areas (PFG, PF, AIP, VIP) connect with the ventral premotor cortex (Marconi et al., 2001; Tanné-Gariépy et al., 2002; Rozzi et al., 2006; Borra et al., 2008; Gamberini et al., 2009; Bakola et al., 2010, 2017; Passarelli et al., 2011, 2018). Input to LIP (Blatt et al., 1990; Lewis and Van Essen, 2000) and PGm (Cavada and Goldman-Rakic, 1989; Passarelli et al., 2018) originates mainly in the oculomotor-related frontal eye fields (FEFs). Segregation of motor projections is not in absolute terms, though, since each parietal area usually receives convergent input from other structures; e.g., PEip receives additional projections from ventral premotor cortex (Tanné-Gariépy et al., 2002; Bakola et al., 2017).

A relative segregation of sensory-specific projections has been described along the rostral-caudal dimension, with somatic-related input targeting heavily rostral parietal areas (Rozzi et al., 2006; Bakola et al., 2013; Padberg et al., 2019). Visual inputs (in particular representations of peripheral vision) are prominent in caudal parietal areas, however there is variation in the source of visual afferents to PPC. For example, numerous afferents to V6A (Passarelli et al., 2011) and LIP (Lewis and Van Essen, 2000) originate in area V6, whereas caudal inferior parietal lobe receives almost exclusively projections from the motion area MST of the temporal cortex (Rozzi et al., 2006). Several projections to MIP and PGm arrive also from the putative visual region

FIGURE 1 | Examples of the distribution of multiple movement preparatory/execution signals related to different effectors in posterior parietal areas. Only the most frequent movement types are illustrated. Mouth-related actions refer mainly to biting. Data derived from electrophysiological studies: Ferraina et al. (1997) (PGm); Battaglia-Mayer et al. (2001) (PEc); Dickinson et al. (2003) (LIP); Kutz et al. (2003); Fattori et al. (2005, 2010) (V6A); Gardner et al. (2007) (PEip); Rozzi et al. (2008) (PG, PFG, PF); Archambault et al. (2009) (PE). Lateral (bottom) and medial (top) views of a macaque brain show our current knowledge of the anatomical organization of posterior parietal cortex (PPC). Abbreviations of sulci: cgs, cingulate; cs, central; ips, intraparietal; lf, lateral fissure; ls, luneate; sts, superior temporal. D, dorsal; R, rostral.

(Kobayashi and Amaral, 2003), ventral to PGm (Bakola et al., 2017; Passarelli et al., 2018). In addition to sensorimotor input, PPC receives segregated input from other systems. For example, caudal/medial areas receive projections from limbic fields of the brain (Rozzi et al., 2006; Bakola et al., 2017; Passarelli et al., 2018). These include projections from the posterior cingulate and retrosplenial regions and area prostriata (Yu et al., 2012) and likely represent routes by which information about spatial orientation and memory reaches parts of PPC (Vann et al., 2009; Kravitz et al., 2011).

Despite the diversity of extrinsic connections, short-range intrinsic connections between adjacent parietal areas form a substantial component of areal connectivity, highlighting the potentially large influence of local processing in defining the function of PPC sectors (Caminiti et al., 2017). This organization may support synergistic actions of different effectors to produce meaningful movements (Kaas and Stepniewska, 2016; Catani et al., 2017).

FUNCTIONAL RESPONSE PROPERTIES IN INDIVIDUAL REGIONS OF THE POSTERIOR PARIETAL CORTEX

Two exemplar nodes of the functional specialization view on PPC are areas AIP and LIP that have been associated with the control of hand-object interactions required for grasping and for the guidance of eye movements, respectively (Gallese et al., 1994; Andersen et al., 1998; Murata et al., 2000; Cui and Andersen, 2007). By comparison, planning and execution of reaching movements appear to be distributed in several areas of the superior (V6A, PEc, MIP and PE/PEip) and inferior parietal

lobe (Snyder et al., 1997; Battaglia-Mayer et al., 2000, 2007; Fattori et al., 2005; Heider et al., 2010; McGuire and Sabes, 2011; Hadjidimitrakis et al., 2012, 2015).

Influential models for parallel parietal-frontal networks for motor actions have dominated parietal research in the past (Jeannerod et al., 1995; Matelli and Luppino, 2001). Accordingly, reach-related signals flow from the superior parietal to the dorsal premotor cortex and grasp-related activity is conveyed from AIP to ventral premotor cortex; both streams converge to the primary motor cortex (Burman et al., 2014; Dea et al., 2016). Re-evaluation of these models became necessary after studies showing that individual premotor neurons carried both reaching and grasping information (Raos et al., 2004; Stark et al., 2007). Along these lines, later work reported grasping parameters to be coded in the traditionally reaching domains of the superior parietal cortex (Chen et al., 2009; Fattori et al., 2010). Furthermore, single AIP neurons encoded both the reaching direction and grip type (Lehmann and Scherberger, 2013).

Additional evidence for the mixing of neural signals comes from work on the spatial reference frames used for reaching movements. Until recently, the dominant view was that neurons in each parietal area have uniform reference frames. A serial organization of reach-related responses along the extent of PPC has been reported, with responses coding target locations relative to the eyes (eye-centered frame) recorded caudally and responses coding locations in head-, body- and hand-centered frame rostrally (Flanders et al., 1992). This view found support in studies that showed eye-centered reference frames caudally in the parietal reach region (PRR, Snyder et al., 1997) and hand-centered representations rostrally in area PE (Lacquaniti et al., 1995; Batista et al., 1999; Buneo et al., 2002; Marzocchi et al., 2008). However, later work showed that neurons in single PPC areas encode reaches relative to the eye, hand, head and body (Mullette-Gillman et al., 2009; Chang and Snyder, 2010; McGuire and Sabes, 2011; Hadjidimitrakis et al., 2014b; Bosco et al., 2016; Piserchia et al., 2017). The presence of mixed, eye- and limb-centered, reference frames within several PPC areas challenges the one-to-one association of a particular type of reference frame with one region and, subsequently, the view of serial reference frame transformations across the PPC "reach" network (McGuire and Sabes, 2011).

Mixing of signals has also been observed at another level of movement control. The distance and direction of reach goals, which were considered to have independent neuronal substrates (Crawford et al., 2011), were encoded by largely overlapping neuronal populations in V6A and PEc (Hadjidimitrakis et al., 2014a, 2015; Filippini et al., 2018). Furthermore, PRR neurons can simultaneously encode multiple potential movement goals (Baldauf et al., 2008; Klaes et al., 2011), thus further illustrating the richness of the selectivity.

In a recent human study, Zhang et al. (2017) reported a mixture of effector representations in populations of neurons in the putative homolog of macaque AIP, arguing against a strict anatomical segregation of body parts. Using fMRI repetition suppression, Heed et al. (2016) examined activity in the PPC in humans performing delayed eye, hand and foot movements to visual targets. They reported a gradient of organization schemes along the extent of PPC, with a region activated independently of the effector used among regions showing effector specificity. Accordingly, the view that emerges is that the primate PPC hosts multiple representations of motor actions, with individual areas and networks (e.g., reaching network) showing only a relative emphasis on a particular effector or movement type.

A POTENTIAL NETWORK FOR EYE-ARM COORDINATION

The mixed selectivity and overlapping representations for different movements in PPC make it an ideal site for mediating complex behaviors like eye-hand coordination. Indeed, growing evidence suggests that coordinated behaviors, such as eye-hand movements, rely on parietal circuits. Reaction times for eye and hand movements are correlated (Dean et al., 2011), suggesting a common neural mechanism. The mixing of various types of signals in single PPC neurons and sectors could be interpreted as a manifestation of coordinated activity. For example, most LIP neurons fire stronger when a combined reach and saccade is planned compared to a saccade alone (Hagan et al., 2012). Neural correlates for single and combined eye- and arm-related movements were reported in several PPC fields (Battaglia-Mayer et al., 2001; Calton et al., 2002; Dickinson et al., 2003), with activity being usually weaker for the non-preferred movement. Moreover, neural responses are modulated by static eye and arm position in PEc, V6A and the caudal inferior parietal lobe (Battaglia-Mayer et al., 2000, 2007; Breveglieri et al., 2012, 2014; Piserchia et al., 2017).

The mixing of signals within PPC may result from the short-range intrinsic connections between adjacent parietal areas (Caminiti et al., 2017). In order to understand the mixed selectivity and how it relates to complex behaviors, simultaneous recordings from multiple PPC areas are necessary. However, very few works have employed this method in PPC (e.g., Cui and Andersen, 2007; Dean et al., 2012). By comparison, increasingly interactions between areas of the frontal and parietal cortex are being studied. Multi-area recordings in primates allow for correlations between the activity across areas to be studied and have complemented non-invasive work using fMRI and MEG.

In electrophysiological studies, the local field potential (LFP) has been instrumental in understanding the relationship in neural activity across brain areas. The LFP is composed of synaptic and spiking activity in the vicinity of the recording electrode (Mitzdorf, 1985), and gives an estimate of the population activity. Like spiking-activity, the LFP power is tuned to saccade and reach direction in LIP and PRR, respectively (Pesaran et al., 2002; Scherberger et al., 2005). Synchrony, or coherence, between the firing rates of individual neurons and the LFP at different frequencies may reflect the processing of different types of information (Fries, 2005). During coordinated eye-hand movements, the beta-band (∼15–30 Hz) LFP activity decreases around movement initiation in both LIP and PRR, and correlates with the reaction times for coordinated reach and saccades (but not for saccades made alone, Dean et al., 2012). Furthermore, LIP neurons with reduced activity during eye-hand

movements, compared to saccades, tend to be coherent with the beta-band LFP (Hagan et al., 2012) and their firing rate predicts the reaction times of coordinated eye-hand movements. This suggests that these neurons participate in a neural circuit that orchestrates coordinated eye-hand movements (Dean et al., 2012). Coherent activity across areas may also contribute to the processing of cognitive signals such as decision-making (Hawellek et al., 2016; Wong et al., 2016) and visual attention in PPC (Buschman and Miller, 2007; Saalmann et al., 2007; Gregoriou et al., 2009).

The studies of LFP-firing coherence are limited in their ability to provide causal evidence of the role of the PPC in eye-hand coordination. In this regard, a number of inactivation studies in PPC have provided more direct evidence, with two works reporting effects on limb (but not eye) movements (Hwang et al., 2012; Yttri et al., 2014), whereas another one found disrupted eye-hand correlations after bilateral inactivation (Battaglia-Mayer et al., 2013). Furthermore, unilateral inactivation of LIP combined with fMRI resulted in rapid spatial reorganization in the active hemisphere (Wilke et al., 2012), suggesting that the functions of PPC are likely spread over a wider network that extends over both hemispheres. This could also explain recent evidence showing no effect of unilateral LIP inactivation on decision-making (Katz et al., 2016). Similarly, inactivation of VIP had no effect on behavior in a heading discrimination task (Chen et al., 2016). In humans, fMRI-guided transcranial magnetic stimulation demonstrated a causal role of the anterior portion of the IPS to reaching (Reichenbach et al., 2011). Overall, inactivation evidence should be treated cautiously. More sensitive activity manipulations could be useful to determine how PPC nodes contribute to motor behaviors. The use of sophisticated tools such as optogenetics in primates (Jazayeri et al., 2012; Watakabe et al., 2016; El-Shamayleh et al., 2017) could help overcome current limitations.

IMPLICATIONS OF MIXED SELECTIVITY IN THE PPC FOR MEDICAL INTERVENTIONS

The diversity of signals within the PPC has sparked great interest to the neuroprosthetic community. For patients suffering from loss of function due to paralysis or amputation of a limb, there can be great difficulty in interacting with people or everyday objects. Brain machine interfaces (BMIs) offer some hope in helping remedy these difficulties. A BMI is a device that can record neural activity from the brain while subjects think about a certain task, and then *via* a decoder, extract the subject's intentions. These decoded intentions are used to control external devices that can vary from a cursor on a monitor, to an anthromorphic robotic arm and hand, to a functional electrical stimulator to activate paralyzed muscles.

Most commonly, electrodes are implanted in the primary motor and premotor areas while patients use motor imagery to provide the necessary input to these BMIs (Markowitz et al., 2011; Hochberg et al., 2012; Collinger et al., 2013). Devices implanted in the motor areas typically decode the trajectory of an effector. Early studies showed that PPC neurons

could be used in conjunction with frontal motor areas to control closed loop BMIs, however it was unclear to what extent the PPC neurons contributed to the efficacy of these devices (Wessberg et al., 2000). In a study that compared offline decoding of hand position and velocity in non-human primates, decoding with PPC neurons was inferior to the decoding performance achieved with primary motor and dorsal premotor cortex (Carmena et al., 2003), possibly indicating that the PPC neurons were not contributing much to the overall control.

However, Musallam et al. (2004) went on to demonstrate that high level movement goal information as well as expected reward values of different targets could be decoded from signals in PRR to control a cursor on the screen during a BMI task. These control signals could be generated in the absence of an actual movement. The goal signals allow an abstraction away from the low-level commands necessary to achieve the wanted action as well as the device that actually enacts the action. These low-level commands can be generated through external optimal control algorithms. Goals for multiple sequential movements are planned in PRR (Baldauf et al., 2008) but not in the superior parietal convexity (Li and Cui, 2013) providing a rich mix of signals.

However, soon after this, trajectory information was successfully decoded from the medial bank of the IPS as well as the dorsal convexity to allow control of a 2-dimensional (2D; Mulliken et al., 2008a,b) as well as 3D (Hauschild et al., 2012) cursor on a screen. Decoding algorithms to incorporate the cognitive neural signals and the trajectory information will also provide increased performance compared to each type of signal alone (Shanechi et al., 2013a,b). These studies primarily focused on decoding of spiking activity, but similar information could be extracted from the LFP (Andersen et al., 2004; Scherberger et al., 2005).

The clinical relevance of the PPC to neural prosthetics was demonstrated in the first human trial of a BMI that utilized neural signals from the PPC (Aflalo et al., 2015). In this study, a tetraplegic patient was implanted with electrode arrays in putative areas 5d/PE and AIP and could successfully control 2D and 3D cursors as well as a robotic limb. Therefore, exploiting the richness of information in the PPC may be an advantageous strategy for developing more efficient BMIs.

CONCLUDING REMARKS

Despite decades of research, a definitive understanding of how individual brains areas are defined, perform distinct computations, and interact with other brain areas remains elusive. The PPC has proved an ideal test bed for understanding how the underlying neural architecture supports a range of sensory, motor and cognitive functions. Anatomy and physiology provide distinct lines of evidence for characterizing the brain areas of the PPC less as a cluster of finite regions and more as a network of integrated areas that may flexibly form the neural basis for diverse functions. The future of systems neuroscience is in understanding how these brain areas work in concert with one another and how the neural dynamics can be used for powering the next generation of prosthetic devices.

AUTHOR CONTRIBUTIONS

KH, SB, YW and MH contributed to the preparation, writing and revising of this text.

FUNDING

ACKNOWLEDGMENTS

REFERENCES

Aflalo, T., Kellis, S., Klaes, C., Lee, B., Shi, Y., Pejsa, K., et al. (2015). Decoding motor imagery from the posterior parietal cortex of a tetraplegic human. *Science* 348, 906–910. doi: 10.1126/science.aaa5417

Andersen, R. A., Andersen, K. N., Hwang, E. J., and Hauschild, M. (2014). Optic ataxia: from Balint's syndrome to the parietal reach region. *Neuron* 81, 967–983. doi: 10.1016/j.neuron.2014.02.025

Andersen, R. A., and Cui, H. (2009). Intention, action planning, and decision making in parietal-frontal circuits. *Neuron* 63, 568–583. doi: 10.1016/j.neuron.2009.08.028

Andersen, R. A., Musallam, S., and Pesaran, B. (2004). Selecting the signals for a brain-machine interface. *Curr. Opin. Neurobiol.* 14, 720–726. doi: 10.1016/j.conb.2004.10.005

Andersen, R. A., Snyder, L. H., Batista, A. P., Buneo, C. A., and Cohen, Y. E. (1998). Posterior parietal areas specialized for eye movements (LIP) and reach (PRR) using a common coordinate frame. *Novartis Found. Symp.* 218, 109–122; discussion 122–128, 171–175. doi: 10.1002/9780470515563.ch7

Arcaro, M. J., Pinsk, M. A., Li, X., and Kastner, S. (2011). Visuotopic organization of macaque posterior parietal cortex: a functional magnetic resonance imaging study. *J. Neurosci.* 31, 2064–2078. doi: 10.1523/JNEUROSCI.3334-10.2011

Archambault, P. S., Caminiti, R., and Battaglia-Mayer, A. (2009). Cortical mechanisms for online control of hand movement trajectory: the role of the posterior parietal cortex. *Cereb. Cortex* 19, 2848–2864. doi: 10.1093/cercor/bhp058

Bakola, S., Gamberini, M., Passarelli, L., Fattori, P., and Galletti, C. (2010). Cortical connections of parietal field PEc in the macaque: linking vision and somatic sensation for the control of limb action. *Cereb. Cortex* 20, 2592–2604. doi: 10.1093/cercor/bhq007

Bakola, S., Passarelli, L., Gamberini, M., Fattori, P., and Galletti, C. (2013). Cortical connectivity suggests a role in limb coordination for macaque area PE of the superior parietal cortex. *J. Neurosci.* 33, 6648–6658. doi: 10.1523/JNEUROSCI.4685-12.2013

Bakola, S., Passarelli, L., Huynh, T., Impieri, D., Worthy, K. H., Fattori, P., et al. (2017). Cortical afferents and myeloarchitecture distinguish the medial intraparietal area (MIP) from neighboring subdivisions of the macaque cortex. *eNeuro* 4:ENEURO.0344-17.2017. doi: 10.1523/eneuro.0344-17.2017

Baldauf, D., Cui, H., and Andersen, R. A. (2008). The posterior parietal cortex encodes in parallel both goals for double-reach sequences. *J. Neurosci.* 28, 10081–10089. doi: 10.1523/JNEUROSCI.3423-08.2008

Batista, A. P., Buneo, C. A., Snyder, L. H., and Andersen, R. A. (1999). Reach plans in eye-centered coordinates. *Science* 285, 257–260. doi: 10.1126/science.285.5425.257

Battaglia-Mayer, A., Ferraina, S., Genovesio, A., Marconi, B., Squatrito, S., Molinari, M., et al. (2001). Eye-hand coordination during reaching. II. An analysis of the relationships between visuomanual signals in parietal cortex and parieto-frontal association projections. *Cereb. Cortex* 11, 528–544. doi: 10.1093/cercor/11.6.528

Battaglia-Mayer, A., Ferraina, S., Mitsuda, T., Marconi, B., Genovesio, A., Onorati, P., et al. (2000). Early coding or reaching in the parietooccipital cortex. *J. Neurophysiol.* 83, 2374–2391. doi: 10.1152/jn.2000.83.4.2374

Battaglia-Mayer, A., Ferrari-Toniolo, S., Visco-Comandini, F., Archambault, P. S., Saberi-Moghadam, S., and Caminiti, R. (2013). Impairment of online control of hand andD eye movements in a monkey model of optic ataxia. *Cereb. Cortex* 23, 2644–2656. doi: 10.1093/cercor/bhs250

Battaglia-Mayer, A., Mascaro, M., and Caminiti, R. (2007). Temporal evolution and strength of neural activity in parietal cortex during eye and hand movements. *Cereb. Cortex* 17, 1350–1363. doi: 10.1093/cercor/bhl046

Blatt, G. J., Andersen, R. A., and Stoner, G. R. (1990). Visual receptive field organization and cortico-cortical connections of the lateral intraparietal area (area LIP) in the macaque. *J. Comp. Neurol.* 299, 421–445. doi: 10.1002/cne.902990404

Borra, E., Belmalih, A., Calzavara, R., Gerbella, M., Murata, A., Rozzi, S., et al. (2008). Cortical connections of the macaque anterior intraparietal (AIP) area. *Cereb. Cortex* 18, 1094–1111. doi: 10.1093/cercor/bhm146

Bosco, A., Breveglieri, R., Hadjidimitrakis, K., Galletti, C., and Fattori, P. (2016). Reference frames for reaching when decoupling eye and target position in depth and direction. *Sci. Rep.* 6:21646. doi: 10.1038/srep21646

Breveglieri, R., Galletti, C., Dal Bò, G., Hadjidimitrakis, K., and Fattori, P. (2014). Multiple aspects of neural activity during reaching preparation in the medial posterior parietal area V6A. *J. Cogn. Neurosci.* 26, 878–895. doi: 10.1162/jocn_a_00510

Breveglieri, R., Hadjidimitrakis, K., Bosco, A., Sabatini, S. P., Galletti, C., and Fattori, P. (2012). Eye position encoding in three-dimensional space: integration of version and vergence signals in the medial posterior parietal cortex. *J. Neurosci.* 32, 159–169. doi: 10.1523/JNEUROSCI.4028-11.2012

Buneo, C. A., Jarvis, M. R., Batista, A. P., and Andersen, R. A. (2002). Direct visuomotor transformations for reaching. *Nature* 416, 632–636. doi: 10.1038/416632a

Burman, K. J., Bakola, S., Richardson, K. E., Reser, D. H., and Rosa, M. G. P. (2014). Patterns of cortical input to the primary motor area in the marmoset monkey. *J. Comp. Neurol.* 522, 811–843. doi: 10.1002/cne.23447

Buschman, T. J., and Miller, E. K. (2007). Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. *Science* 315, 1860–1862. doi: 10.1126/science.1138071

Calton, J. L., Dickinson, A. R., and Snyder, L. H. (2002). Non-spatial, motor-specific activation in posterior parietal cortex. *Nat. Neurosci.* 5, 580–588. doi: 10.1038/nn0602-862

Caminiti, R., Borra, E., Visco-Comandini, F., Battaglia-Mayer, A., Averbeck, B. B., and Luppino, G. (2017). Computational architecture of the parieto-frontal network underlying cognitive-motor control in monkeys. *eNeuro* 4:ENEURO.0306-16.2017. doi: 10.1523/eneuro.0306-16.2017

Carmena, J. M., Lebedev, M. A., Crist, R. E., O'Doherty, J. E., Santucci, D. M., Dimitrov, D. F., et al. (2003). Learning to control a brain-machine interface for reaching and grasping by primates. *PLoS Biol.* 1:e42. doi: 10.1371/journal.pbio.0000042

Catani, M., Robertsson, N., Beyh, A., Huynh, V., de Santiago Requejo, F., Howells, H., et al. (2017). Short parietal lobe connections of the human and monkey brain. *Cortex* 97, 339–357. doi: 10.1016/j.cortex.2017.10.022

Cavada, C., and Goldman-Rakic, P. S. (1989). Posterior parietal cortex in rhesus monkey: II. Evidence for segregated corticocortical networks linking sensory and limbic areas with the frontal lobe. *J. Comp. Neurol.* 287, 422–445. doi: 10.1002/cne.902870403

Chang, S. W. C., and Snyder, L. H. (2010). Idiosyncratic and systematic aspects of spatial representations in the macaque parietal cortex. *Proc. Natl. Acad. Sci. U S A* 107, 7951–7956. doi: 10.1073/pnas.0913209107

Chen, A., Gu, Y., Liu, S., DeAngelis, G. C., and Angelaki, D. E. (2016). Evidence for a causal contribution of macaque vestibular, but not intraparietal, cortex to heading perception. *J. Neurosci.* 36, 3789–3798. doi: 10.1523/JNEUROSCI.2485-15.2016

Chen, J., Reitzen, S. D., Kohlenstein, J. B., and Gardner, E. P. (2009). Neural representation of hand kinematics during prehension in posterior parietal cortex of the macaque monkey. *J. Neurophysiol.* 102, 3310–3328. doi: 10.1152/jn.90942.2008

Collinger, J. L., Wodlinger, B., Downey, J. E., Wang, W., Tyler-Kabara, E. C., Weber, D. J., et al. (2013). High-performance neuroprosthetic control by an individual with tetraplegia. *Lancet* 381, 557–564. doi: 10.1016/S0140-6736(12)61816-9

Crawford, J. D., Henriques, D. Y., and Medendorp, W. P. (2011). Three-dimensional transformations for goal-directed action. *Annu. Rev. Neurosci.* 34, 309–331. doi: 10.1146/annurev-neuro-061010-113749

Cui, H., and Andersen, R. A. (2007). Posterior parietal cortex encodes autonomously selected motor plans. *Neuron* 56, 552–559. doi: 10.1016/j.neuron.2007.09.031

Dea, M., Hamadjida, A., Elgbeili, G., Quessy, S., and Dancause, N. (2016). Different patterns of cortical inputs to subregions of the primary motor cortex hand representation in cebus apella. *Cereb. Cortex* 26, 1747–1761. doi: 10.1093/cercor/bhv324

Dean, H. L., Hagan, M. A., and Pesaran, B. (2012). Only coherent spiking in posterior parietal cortex coordinates looking and reaching. *Neuron* 73, 829–841. doi: 10.1016/j.neuron.2011.12.035

Dean, H. L., Martí, D., Tsui, E., Rinzel, J., and Pesaran, B. (2011). Reaction time correlations during eye-hand coordination: behavior and modeling. *J. Neurosci.* 31, 2399–2412. doi: 10.1523/JNEUROSCI.4591-10.2011

Dickinson, A. R., Calton, J. L., and Snyder, L. H. (2003). Nonspatial saccade-specific activation in area LIP of monkey parietal cortex. *J. Neurophysiol.* 90, 2460–2464. doi: 10.1152/jn.00788.2002

El-Shamayleh, Y., Kojima, Y., Soetedjo, R., and Horwitz, G. D. (2017). Selective optogenetic control of purkinje cells in monkey cerebellum. *Neuron* 95, 51.e4–62.e4. doi: 10.1016/j.neuron.2017.06.002

Fattori, P., Kutz, D. F., Breveglieri, R., Marzocchi, N., and Galletti, C. (2005). Spatial tuning of reaching activity in the medial parieto-occipital cortex (area V6A) of macaque monkey. *Eur. J. Neurosci.* 22, 956–972. doi: 10.1111/j.1460-9568.2005.04288.x

Fattori, P., Raos, V., Breveglieri, R., Bosco, A., Marzocchi, N., and Galletti, C. (2010). The dorsomedial pathway is not just for reaching: grasping neurons in the medial parieto-occipital cortex of the macaque monkey. *J. Neurosci.* 30, 342–349. doi: 10.1523/JNEUROSCI.3800-09.2010

Ferraina, S., Garasto, M. R., Battaglia-Mayer, A., Ferraresi, P., Johnson, P. B., Lacquaniti, F., et al. (1997). Visual control of hand reaching movement: activity in parietal area 7m. *Eur. J. Neurosci.* 9, 1090–1095. doi: 10.1111/j.1460-9568.1997.tb01460.x

Filippini, M., Breveglieri, R., Hadjidimitrakis, K., Bosco, A., and Fattori, P. (2018). Prediction of reach goals in depth and direction from the parietal cortex. *Cell Rep.* 23, 725–732. doi: 10.1016/j.celrep.2018.03.090

Flanders, M., Helms Tillery, S. I., and Soechting, J. F. (1992). Early stages in a sensorimotor transformation. *Behav. Brain Sci.* 15, 309–362. doi: 10.1017/s0140525x00068813

Fries, P. (2005). A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends Cogn. Sci.* 9, 474–480. doi: 10.1016/j.tics.2005.08.011

Gallese, V., Murata, A., Kaseda, M., Niki, N., and Sakata, H. (1994). Deficit of hand preshaping after muscimol injection in monkey parietal cortex. *Neuroreport* 5, 1525–1529. doi: 10.1097/00001756-199407000-00029

Gamberini, M., Passarelli, L., Fattori, P., Zucchelli, M., Bakola, S., Luppino, G., et al. (2009). Cortical connections of the visuomotor parietooccipital area V6Ad of the macaque monkey. *J. Comp. Neurol.* 513, 622–642. doi: 10.1002/cne.21980

Gardner, E. P., Babu, K. S., Reitzen, S. D., Ghosh, S., Brown, A. S., Chen, J., et al. (2007). Neurophysiology of prehension. I. Posterior parietal cortex and object-oriented hand behaviors. *J. Neurophysiol.* 97, 387–406. doi: 10.1152/jn.00558.2006

Gregoriou, G. G., Gotts, S. J., Zhou, H., and Desimone, R. (2009). High-frequency, long-range coupling between prefrontal and visual cortex during attention. *Science* 324, 1207–1210. doi: 10.1126/science.1171402

Hadjidimitrakis, K., Bertozzi, F., Breveglieri, R., Bosco, A., Galletti, C., and Fattori, P. (2014a). Common neural substrate for processing depth and direction signals for reaching in the monkey medial posterior parietal cortex. *Cereb. Cortex* 24, 1645–1657. doi: 10.1093/cercor/bht021

Hadjidimitrakis, K., Bertozzi, F., Breveglieri, R., Fattori, P., and Galletti, C. (2014b). Body-centered, mixed, but not hand-centered coding of visual targets in the medial posterior parietal cortex during reaches in 3D space. *Cereb. Cortex* 24, 3209–3220. doi: 10.1093/cercor/bht181

Hadjidimitrakis, K., Breveglieri, R., Bosco, A., and Fattori, P. (2012). Three-dimensional eye position signals shape both peripersonal space and arm movement activity in the medial posterior parietal cortex. *Front. Integr. Neurosci.* 6:37. doi: 10.3389/fnint.2012.00037

Hadjidimitrakis, K., Dal Bo', G., Breveglieri, R., Galletti, C., and Fattori, P. (2015). Overlapping representations for reach depth and direction in caudal superior parietal lobule of macaques. *J. Neurophysiol.* 114, 2340–2352. doi: 10.1152/jn.00486.2015

Hagan, M. A., Dean, H. L., and Pesaran, B. (2012). Spike-field activity in parietal area LIP during coordinated reach and saccade movements. *J. Neurophysiol.* 107, 1275–1290. doi: 10.1152/jn.00867.2011

Hauschild, M., Mulliken, G. H., Fineman, I., Loeb, G. E., and Andersen, R. A. (2012). Cognitive signals for brain-machine interfaces in posterior parietal cortex include continuous 3D trajectory commands. *Proc. Natl. Acad. Sci. U S A* 109, 17075–17080. doi: 10.1073/pnas.1215092109

Hawellek, D. J., Wong, Y. T., and Pesaran, B. (2016). Temporal coding of reward-guided choice in the posterior parietal cortex. *Proc. Natl. Acad. Sci. U S A* 113, 13492–13497. doi: 10.1073/pnas.1606479113

Hayhoe, M. M., Shrivastava, A., Mruczek, R., and Pelz, J. B. (2003). Visual memory and motor planning in a natural task. *J. Vis.* 3, 49–63. doi: 10.1167/3.1.6

Heed, T., Leone, F. T., Toni, I., and Medendorp, W. P. (2016). Functional versus effector-specific organization of the human posterior parietal cortex: revisited. *J. Neurophysiol.* 116, 1885–1899. doi: 10.1152/jn.00312.2014

Heider, B., Karnik, A., Ramalingam, N., and Siegel, R. M. (2010). Neural representation during visually guided reaching in macaque posterior parietal cortex. *J. Neurophysiol.* 104, 3494–3509. doi: 10.1152/jn.010 50.2009

Hochberg, L. R., Bacher, D., Jarosiewicz, B., Masse, N. Y., Simeral, J. D., Vogel, J., et al. (2012). Reach and grasp by people with tetraplegia using a neurally controlled robotic arm. *Nature* 485, 372–375. doi: 10.1038/nature 11076

Husain, M., and Nachev, P. (2007). Space and the parietal cortex. *Trends Cogn. Sci.* 11, 30–36. doi: 10.1016/j.tics.2006.10.011

Hwang, E. J., Hauschild, M., Wilke, M., and Andersen, R. A. (2012). Inactivation of the parietal reach region causes optic ataxia, impairing reaches but not saccades. *Neuron* 76, 1021–1029. doi: 10.1016/j.neuron.2012.10.030

Jazayeri, M., Lindbloom-Brown, Z., and Horwitz, G. D. (2012). Saccadic eye movements evoked by optogenetic activation of primate V1. *Nat. Neurosci.* 15, 1368–1370. doi: 10.1038/nn.3210

Jeannerod, M., Arbib, M. A., Rizzolatti, G., and Sakata, H. (1995). Grasping objects: the cortical mechanisms of visuomotor transformation. *Trends Neurosci.* 18, 314–320. doi: 10.1016/0166-2236(95)93921-j

Johansson, R. S., Westling, G., Bäckström, A., and Flanagan, J. R. (2001). Eye-hand coordination in object manipulation. *J. Neurosci.* 21, 6917–6932. doi: 10.1523/JNEUROSCI.21-17-06917.2001

Kaas, J. H., and Stepniewska, I. (2016). Evolution of posterior parietal cortex and parietal-frontal networks for specific actions in primates. *J. Comp. Neurol.* 524, 595–608. doi: 10.1002/cne.23838

Katz, L. N., Yates, J. L., Pillow, J. W., and Huk, A. C. (2016). Dissociated functional significance of decision-related activity in the primate dorsal stream. *Nature* 535, 285–288. doi: 10.1038/nature18617

Klaes, C., Westendorff, S., Chakrabarti, S., and Gail, A. (2011). Choosing goals, not rules: deciding among rule-based action plans. *Neuron* 70, 536–548. doi: 10.1016/j.neuron.2011.02.053

Kobayashi, Y., and Amaral, D. G. (2003). Macaque monkey retrosplenial cortex: II. cortical afferents. *J. Comp. Neurol.* 466, 48–79. doi: 10.1002/cne.10883

Kravitz, D. J., Saleem, K. S., Baker, C. I., and Mishkin, M. (2011). A new neural framework for visuospatial processing. *Nat. Rev. Neurosci.* 12, 217–230. doi: 10.1038/nrn3008

Kutz, D. F., Fattori, P., Gamberini, M., Breveglieri, R., and Galletti, C. (2003). Early- and late-responding cells to saccadic eye movements in the cortical area V6A of macaque monkey. *Exp. Brain Res.* 149, 83–95. doi: 10.1007/s00221-002-1337-9

Lacquaniti, F., Guigon, E., Bianchi, L., Ferraina, S., and Caminiti, R. (1995). Representing spatial information for limb movement: role of area 5 in the monkey. *Cereb. Cortex* 5, 391–409. doi: 10.1093/cercor/5.5.391

Land, M. F., and Hayhoe, M. (2001). In what ways do eye movements contribute to everyday activities? *Vision Res.* 41, 3559–3565. doi: 10.1016/s0042-6989(01)00102-x

Lehmann, S. J., and Scherberger, H. (2013). Reach and gaze representations in macaque parietal and premotor grasp areas. *J. Neurosci.* 33, 7038–7049. doi: 10.1523/JNEUROSCI.5568-12.2013

Lewis, J. W., and Van Essen, D. C. (2000). Corticocortical connections of visual, sensorimotor, and multimodal processing area in the parietal lobe of the macaque monkey. *J. Comp. Neurol.* 428, 112–137. doi: 10.1002/1096-9861(20001204)428:1<112::aid-cne8>3.0.co;2-9

Li, Y., and Cui, H. (2013). Dorsal parietal area 5 encodes immediate reach in sequential arm movements. *J. Neurosci.* 33, 14455–14465. doi: 10.1523/JNEUROSCI.1162-13.2013

Marconi, B., Genovesio, A., Battaglia-Mayer, A., Ferraina, S., Squatrito, S., Molinari, M., et al. (2001). Eye-hand coordination during reaching. I. Anatomical relationships between parietal and frontal cortex. *Cereb. Cortex* 11, 513–527. doi: 10.1093/cercor/11.6.513

Markowitz, D. A., Wong, Y. T., Gray, C. M., and Pesaran, B. (2011). Optimizing the decoding of movement goals from local field potentials in macaque cortex. *J. Neurosci.* 31, 18412–18422. doi: 10.1523/JNEUROSCI.4165-11.2011

Marzocchi, N., Breveglieri, R., Galletti, C., and Fattori, P. (2008). Reaching activity in parietal area V6A of macaque: eye influence on arm activity or retinocentric coding of reaching movements? *Eur. J. Neurosci.* 27, 775–789. doi: 10.1111/j.1460-9568.2008.06021.x

Matelli, M., and Luppino, G. (2001). Parietofrontal circuits for action and space perception in the macaque monkey. *Neuroimage* 14, S27–S32. doi: 10.1006/nimg.2001.0835

McGuire, L. M., and Sabes, P. N. (2011). Heterogeneous representations in the superior parietal lobule are common across reaches to visual and proprioceptive targets. *J. Neurosci.* 31, 6661–6673. doi: 10.1523/jneurosci.2921-10.2011

Mitzdorf, U. (1985). Current source-density method and application in cat cerebral cortex: investigation of evoked potentials and EEG phenomena. *Physiol. Rev.* 65, 37–100. doi: 10.1152/physrev.1985.65.1.37

Mountcastle, V. B., Lynch, J. C., Georgopoulos, A., Sakata, H., and Acuna, C. (1975). Posterior parietal association cortex of the monkey: command functions for operations within extrapersonal space. *J. Neurophysiol.* 38, 871–908. doi: 10.1152/jn.1975.38.4.871

Mullette-Gillman, O. D. A., Cohen, Y. E., and Groh, J. M. (2009). Motor-related signals in the intraparietal cortex encode locations in a hybrid, rather than eye-centered reference frame. *Cereb. Cortex* 19, 1761–1775. doi: 10.1093/cercor/bhn207

Mulliken, G. H., Musallam, S., and Andersen, R. A. (2008a). Decoding trajectories from posterior parietal cortex ensembles. *J. Neurosci.* 28, 12913–12926. doi: 10.1523/jneurosci.1463-08.2008

Mulliken, G. H., Musallam, S., and Andersen, R. A. (2008b). Forward estimation of movement state in posterior parietal cortex. *Proc. Natl. Acad. Sci. U S A* 105, 8170–8177. doi: 10.1073/pnas.0802602105

Murata, A., Gallese, V., Luppino, G., Kaseda, M., and Sakata, H. (2000). Selectivity for the shape, size and orientation of objects for grasping in neurons of monkey parietal area AIP. *J. Neurophysiol.* 83, 2580–2601. doi: 10.1152/jn.2000.83.5.2580

Musallam, S., Corneil, B. D., Greger, B., Scherberger, H., and Andersen, R. A. (2004). Cognitive control signals for neural prosthetics. *Science* 305, 258–262. doi: 10.1126/science.1097938

Padberg, J., Cooke, D. F., Cerkevich, C. M., Kaas, J. H., and Krubitzer, L. (2019). Cortical connections of area 2 and posterior parietal area 5 in macaque monkeys. *J. Comp. Neurol.* 527, 718–737. doi: 10.1002/cne.24453

Passarelli, L., Rosa, M. G. P., Bakola, S., Gamberini, M., Worthy, K. H., Fattori, P., et al. (2018). Uniformity and diversity of cortical projections to precuneate areas in the macaque monkey: what defines area PGm? *Cereb. Cortex* 28, 1700–1717. doi: 10.1093/cercor/bhx067

Passarelli, L., Rosa, M. G. P., Gamberini, M., Bakola, S., Burman, K. J., Fattori, P., et al. (2011). Cortical connections of area V6Av in the macaque: a visual-input node to the eye/hand coordination system. *J. Neurosci.* 31, 1790–1801. doi: 10.1523/jneurosci.4784-10.2011

Pesaran, B., Pezaris, J. S., Sahani, M., Mitra, P. P., and Andersen, R. A. (2002). Temporal structure in neuronal activity during working memory in macaque parietal cortex. *Nat. Neurosci.* 5, 805–811. doi: 10.1038/nn890

Piserchia, V., Breveglieri, R., Hadjidimitrakis, K., Bertozzi, F., Galletti, C., and Fattori, P. (2017). Mixed body/hand reference frame for reaching in 3D space in macaque parietal area PEc. *Cereb. Cortex* 27, 1976–1990. doi: 10.1093/cercor/bhw039

Raos, V., Umiltá, M. A., Gallese, V., and Fogassi, L. (2004). Functional properties of grasping-related neurons in the dorsal premotor area F2 of the macaque monkey. *J. Neurophysiol.* 92, 1990–2002. doi: 10.1152/jn.00154.2004

Reichenbach, A., Bresciani, J. P., Peer, A., Bülthoff, H. H., and Thielscher, A. (2011). Contributions of the PPC to online control of visually guided reaching movements assessed with fMRI-guided TMS. *Cereb. Cortex* 21, 1602–1612. doi: 10.1093/cercor/bhq225

Rizzolatti, G., and Matelli, M. (2003). Two different streams form the dorsal visual system: anatomy and functions. *Exp. Brain Res.* 153, 146–157. doi: 10.1007/s00221-003-1588-0

Rozzi, S., Calzavara, R., Belmalih, A., Borra, E., Gregoriou, G. G., Matelli, M., et al. (2006). Cortical connections of the inferior parietal cortical convexity of the macaque monkey. *Cereb. Cortex* 16, 1389–1417. doi: 10.1093/cercor/bhj076

Rozzi, S., Ferrari, P. F., Bonini, L., Rizzolatti, G., and Fogassi, L. (2008). Functional organization of inferior parietal lobule convexity in the macaque monkey: electrophysiological characterization of motor, sensory and mirror responses and their correlation with cytoarchitectonic areas. *Eur. J. Neurosci.* 28, 1569–1588. doi: 10.1111/j.1460-9568.2008.06395.x

Saalmann, Y. B., Pigarev, I. N., and Vidyasagar, T. R. (2007). Neural mechanisms of visual attention: how top-down feedback highlights relevant locations. *Science* 316, 1612–1615. doi: 10.1126/science.1139140

Savaki, H. E., Gregoriou, G. G., Bakola, S., Raos, V., and Moschovakis, A. K. (2010). The place code of saccade metrics in the lateral bank of the intraparietal sulcus. *J. Neurosci.* 30, 1118–1127. doi: 10.1523/jneurosci.2268-09.2010

Scherberger, H., Jarvis, M. R., and Andersen, R. A. (2005). Cortical local field potential encodes movement intentions in the posterior parietal cortex. *Neuron* 46, 347–354. doi: 10.1016/j.neuron.2005.03.004

Seelke, A. M., Padberg, J. J., Disbrow, E., Purnell, S. M., Recanzone, G., and Krubitzer, L. (2012). Topographic maps within Brodmann's area 5 of macaque monkeys. *Cereb. Cortex* 22, 1834–1850. doi: 10.1093/cercor/bhr257

Shanechi, M. M., Williams, Z. M., Wornell, G. W., Hu, R. C., Powers, M., and Brown, E. N. (2013a). A real-time brain-machine interface combining motor target and trajectory intent using an optimal feedback control design. *PLoS One* 8:e59049. doi: 10.1371/journal.pone.0059049

Shanechi, M. M., Wornell, G. W., Williams, Z. M., and Brown, E. N. (2013b). Feedback-controlled parallel point process filter for estimation of goal-directed movements from neural signals. *IEEE Trans. Neural Syst. Rehabil. Eng.* 21, 129–140. doi: 10.1109/TNSRE.2012.2221743

Snyder, L. H., Batista, A. P., and Andersen, R. A. (1997). Coding of intention in the posterior parietal cortex. *Nature* 386, 167–170. doi: 10.1038/386167a0

Stark, E., Asher, I., and Abeles, M. (2007). Encoding of reach and grasp by single neurons in premotor cortex is independent of recording site. *J. Neurophysiol.* 97, 3351–3364. doi: 10.1152/jn.01328.2006

Tanné-Gariépy, J., Rouiller, E. M., and Boussaoud, D. (2002). Parietal inputs to dorsal versus ventral premotor areas in the macaque monkey: evidence for largely segregated visuomotor pathways. *Exp. Brain Res.* 145, 91–103. doi: 10.1007/s00221-002-1078-9

Vann, S. D., Aggleton, J. P., and Maguire, E. A. (2009). What does the retrosplenial cortex do? *Nat. Rev. Neurosci.* 10, 792–802. doi: 10.1038/nrn2733

Vesia, M., and Crawford, J. D. (2012). Specialization of reach function in human posterior parietal cortex. *Exp. Brain Res.* 221, 1–18. doi: 10.1007/s00221-012-3158-9

Watakabe, A., Sadakane, O., Hata, K., Ohtsuka, M., Takaji, M., and Yamamori, T. (2016). Application of viral vectors to the study of neural connectivities and neural circuits in the marmoset brain. *Dev. Neurobiol.* 77, 354–372. doi: 10.1002/dneu.22459

Wessberg, J., Stambaugh, C. R., Kralik, J. D., Beck, P. D., Laubach, M., Chapin, J. K., et al. (2000). Real-time prediction of hand trajectory by ensembles

of cortical neurons in primates. *Nature* 408, 361–365. doi: 10.1038/350 42582

Wilke, M., Kagan, I., and Andersen, R. A. (2012). Functional imaging reveals rapid reorganization of cortical activity after parietal inactivation in monkeys. *Proc. Natl. Acad. Sci. U S A* 109, 8274–8279. doi: 10.1073/pnas.120 4789109

Wong, Y. T., Fabiszak, M. M., Novikov, Y., Daw, N. D., and Pesaran, B. (2016). Coherent neuronal ensembles are rapidly recruited when making a look-reach decision. *Nat. Neurosci.* 19, 327–334. doi: 10.1038/nn.4210

Yttri, E. A., Wang, C., Liu, Y., and Snyder, L. H. (2014). The parietal reach region is limb specific and not involved in eye-hand coordination. *J. Neurophysiol.* 111, 520–532. doi: 10.1152/jn.00058.2013

Yu, H. H., Chaplin, T. A., Davies, A. J., Verma, R., and Rosa, M. G. P. (2012). A specialized area in limbic cortex for fast analysis of peripheral vision. *Curr. Biol.* 22, 1351–1357. doi: 10.1016/j.cub.2012.05.029

Zhang, C. Y., Aflalo, T., Revechkis, B., Rosario, E. R., Ouellette, D., Pouratian, N., et al. (2017). Partially mixed selectivity in human posterior parietal association cortex. *Neuron* 95, 697.e4–708.e4. doi: 10.1016/j.neuron.2017.06.040

Check for updates

Toward a Biologically Plausible Model of LGN-V1 Pathways Based on Efficient Coding

Yanbo Lian[1], David B. Grayden[1,2], Tatiana Kameneva[1,3], Hamish Meffin[4,5*†] and Anthony N. Burkitt[1*†]

[1] Department of Biomedical Engineering, The University of Melbourne, Melbourne, VIC, Australia, [2] Centre for Neural Engineering, The University of Melbourne, Melbourne, VIC, Australia, [3] Faculty of Science, Engineering and Technology, Swinburne University, Melbourne, VIC, Australia, [4] Department of Optometry and Visual Science, The University of Melbourne, Melbourne, VIC, Australia, [5] National Vision Research Institute, The Australian College of Optometry, Melbourne, VIC, Australia

Increasing evidence supports the hypothesis that the visual system employs a sparse code to represent visual stimuli, where information is encoded in an efficient way by a small population of cells that respond to sensory input at a given time. This includes simple cells in primary visual cortex (V1), which are defined by their linear spatial integration of visual stimuli. Various models of sparse coding have been proposed to explain physiological phenomena observed in simple cells. However, these models have usually made the simplifying assumption that inputs to simple cells already incorporate linear spatial summation. This overlooks the fact that these inputs are known to have strong non-linearities such the separation of ON and OFF pathways, or separation of excitatory and inhibitory neurons. Consequently these models ignore a range of important experimental phenomena that are related to the emergence of linear spatial summation from non-linear inputs, such as segregation of ON and OFF sub-regions of simple cell receptive fields, the push-pull effect of excitation and inhibition, and phase-reversed cortico-thalamic feedback. Here, we demonstrate that a two-layer model of the visual pathway from the lateral geniculate nucleus to V1 that incorporates these biological constraints on the neural circuits and is based on sparse coding can account for the emergence of these experimental phenomena, diverse shapes of receptive fields and contrast invariance of orientation tuning of simple cells when the model is trained on natural images. The model suggests that sparse coding can be implemented by the V1 simple cells using neural circuits with a simple biologically plausible architecture.

Keywords: efficient coding, LGN-V1 pathways, biological plausibility, separated ON and OFF sub-regions, push-pull effect, phase-reversed feedback, receptive fields, contrast invariance

1. INTRODUCTION

In early experimental studies of cat striate cortex, Hubel and Wiesel found two main types of cells: simple cells and complex cells (Hubel and Wiesel, 1959, 1962). Simple cells exhibit linear spatial summation of visual stimuli, while complex cells have significant non-linear behavior. This difference is reflected in receptive field (RF) structures of the two types of cells. Receptive fields (RFs) describe spatial patterns of light and dark regions in the visual field that in combination are effective at driving neural response. They are frequently modeled as linear spatial filters. Simple

cells have a single RF filter, reflecting the linear spatial summation properties, while complex cells pool the output for two or more RF filters in a non-linear fashion.

Over the past decades, some important characteristics of simple cell RF have been observed experimentally (with emphasis on cat and primates, but also ferrets). First, simple cells show a range of selectivity for the orientation of visual stimuli, from highly oriented RFs, which are selective to an optimal orientation, to non-oriented RFs, which are insensitive to orientation. Many RFs of simple cells in V1 are oriented, localized, and bandpass (Hubel and Wiesel, 1962, 1968), while non-orientated RFs are seen in all layers of V1 (Hawken et al., 1988; Chapman and Stryker, 1993). Second, RFs of orientation tuned simple cells can be well-described by two-dimensional Gabor functions (Jones and Palmer, 1987a; Ringach, 2002). In addition, both these studies found some blob-like RFs, which are broadly tuned in orientation. Third, RFs of simple cells have spatially segregated ON and OFF sub-regions (Hubel and Wiesel, 1962; Martinez et al., 2005); i.e., the spatial region that excites the simple cell in response to bright (ON) stimuli is separated from the region that excites the cell in response to dark (OFF) stimuli (left column of **Figure 1**). Fourth, simple cells show push-pull responses; i.e., if one stimulus excites a simple cell, the stimulus with opposite contrast, but same location, will inhibit the simple cell (Jones and Palmer, 1987b; Ferster, 1988; Hirsch et al., 1998; Martinez et al., 2005). One example of the push-pull effect can be seen on the left of **Figure 1** where a simple cell is excited by input from a cell in the lateral geniculate nucleus (LGN) responding to dark spots (an OFF LGN cell) but is effectively inhibited by LGN cells responding a bright spot in the same location (an ON LGN cell). Fifth, feedback from simple cells to LGN cells frequently has a phase-reversed influence compared to the feedforward input (Wang et al., 2006); i.e., where the RF of an ON (OFF) LGN cell is overlapped with the ON (OFF) sub-region of the RF of a simple cell, i.e., feedforward excitation, feedback from the simple cell to the LGN cell is suppressive; where an ON (OFF) LGN cell coincides with the OFF (ON) sub-region of a simple cell RF, i.e., effective feedforward suppression, the feedback is facilitatory. This effect of phase-reversed feedback is also illustrated in **Figure 1**, where the influence from a simple cell to LGN cells is opposite to the influence from LGN cells to the same simple cell. Lastly, the orientation tuning property of simple cells are contrast invariant; i.e., the shape and width of orientation tuning curves remain the same for different stimulus contrasts (Sclar and Freeman, 1982; Skottun et al., 1987; Finn et al., 2007; Priebe, 2016).

On the other hand, insights from computational modeling of V1 cells have also been used to explain experimental data. Sparse coding has been proposed by many researchers as a principle employed by the brain to process sensory information. Olshausen and Field reproduced localized, oriented and spatially bandpass RFs of simple cells based on a *sparse coding* model that aimed to reconstruct the input with minimal average activity of neurons (Olshausen and Field, 1996, 1997). However, the original model failed to generate non-oriented RFs observed in experiments (Ringach, 2002). Subsequently, Olshausen and colleagues found that the sparse coding model can produce RFs that lack strong

FIGURE 1 | Illustration of segregated ON and OFF sub-regions, the push-pull effect, and phase-reversed feedback. ON and OFF LGN cells are spatially located in a 2D region. The colors of magenta and green represent excitatory and inhibitory connections, respectively.
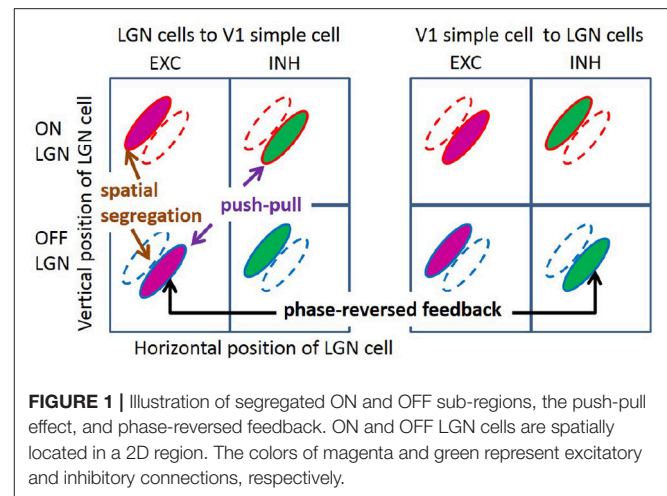
orientation selectivity by having many more model neurons than the number of input image pixels (Olshausen et al., 2009). Rehn and Sommer introduced *hard sparseness* to classical sparse coding, which minimizes the number of active neurons rather than the average activity of neurons in the original model, and demonstrated that the modified sparse coding model can generate diverse shapes of simple cell RFs (Rehn and Sommer, 2007). Zhu and Rozell showed that many visual non-classical RF effects of V1 such as end-stopping, contrast invariance of orientation tuning can emerge from a dynamical system based on sparse coding (Zhu and Rozell, 2013).

These studies were important in explaining the RF structure, but made a number of simplifying assumptions that overlooked many details of biological reality, include some or all of the following. First, the responses of neurons (e.g., firing rates) should be non-negative. Second, the learning rule of synaptic connections should be local where the changes of synaptic efficacy depend only on pre-synaptic and post-synaptic responses. Third, the learning rule should not violate Dale's Law, namely that neurons release the same type of transmitter at all their synapses, and consequently, the synapses are either all excitatory or all inhibitory (Strata and Harvey, 1999). Fourth, the computation of the response of any neuron should be local, such that only neurons synaptically connected to this target neuron can be involved. In addition, a biologically plausible model should also be consistent with important experimental evidence. For LGN-V1 visual pathways, experimental evidence includes the existence of a large amount of cortico-thalamic feedback (Swadlow, 1983; Sherman and Guillery, 1996), long-range excitatory but not inhibitory connections between LGN and V1, and separated ON and OFF channels for LGN input (Hubel and Wiesel, 1962; Ferster et al., 1996; Jin et al., 2008, 2011). The original sparse coding model neglects many of the biological constraints described above.

Several recent studies addressed the issue of biological plausibility by incorporating some of these constraints, while continuing to neglect others. For example, Zylberberg and colleagues designed a spiking network (based on sparse coding)

that can account for diverse shapes of simple cell RFs using lateral inhibition (Zylberberg et al., 2011). The local learning rule and the use of spiking neurons bring some degree of biological plausibility to the model, but the model employs connections that can change sign during learning, which violates Dale's law, and there are not separate channels for ON and OFF LGN input. Additionally, the effect of sparse coding is achieved by competition between units via lateral inhibition, but a recent study suggested that dominant lateral interactions are excitatory in the mouse cortex (Lee et al., 2016). In another modeling work of simple cell RFs, Wiltschut and Hamker designed an efficient coding model with separated ON and OFF LGN cells, and, feedforward, feedback, and lateral connections that can generate various types of simple cell RFs (Wiltschut and Hamker, 2009), but their model does not incorporate Dale's law.

As with earlier studies (Olshausen and Field, 1996, 1997; Rehn and Sommer, 2007; Olshausen et al., 2009), these more recent studies (Wiltschut and Hamker, 2009; Zylberberg et al., 2011), incorporating biological constraints, have continued to focus on the RF structure of simple cells, while largely neglecting the experimental phenomena shown in **Figure 1**. This is because they have typically not separated inputs from ON and OFF LGN cells, which is a key distinction underlying all the phenomena listed in **Figure 1**. One important question in this regard is how these non-linear (half-wave rectified) LGN inputs are combined to give linear RFs for simple cells and whether this causes the experimental phenomena listed in **Figure 1**. To our knowledge, Jehee and Ballard are the only researchers that have explicitly explained the effect of phase-reversed feedback using a model based on predictive coding (Jehee and Ballard, 2009). However, the RFs generated by their model do not match well with those observed in experiments and the push-pull effect for simple cells has not been explained. In addition, the formula for calculating responses of model neurons (Jehee and Ballard, 2009, Equation 7) is not local and the learning rule neglects Dale's law.

In this paper, we propose a two-layer model of LGN-V1 visual pathways that can account for experimental phenomena:

- Segregated ON and OFF sub-regions of simple cells,
- The push-pull effect for simple cells,
- Phase-reversed cortico-thalamic feedback,
- Diverse shapes of RFs (oriented and non-oriented),
- Contrast invariance of orientation tuning.

Our model is biologically plausible by incorporating:

- Separate channels of ON and OFF LGN input,
- Non-negative neural responses,
- Local learning rule,
- Dale's law,
- Local computation,
- Dynamics of rate-based model neurons,
- Feedback from V1 to LGN.

The first layer consists of ON and OFF LGN cells and the second layer consists of simple cells. The connections from the first layer to the second layer (feedforward connections) and from the second layer to the first layer (feedback connections) consist of separate excitatory and inhibitory connections. Even

though the inhibitory connections between LGN and V1 should be implemented via intermediate populations of inhibitory interneurons, we use neurons that have both excitatory and inhibitory connections to simplify the circuit. This aspect of the model is not biologically plausible, but possible biologically plausible neural circuits for implementing inhibitory connections are proposed in the Discussion section. The model presented here is relevant to visual cortices both with and without an orientation columnar organization.

The novelty of the model proposed here is that it models LGN-V1 pathways using segregated ON and OFF LGN channels and separate excitatory and inhibitory connections to investigate the structure of connections between LGN and simple cells to explain a wide range of experimental phenomena. In addition, it can generate a wide variety of experimentally observed RFs of simple cells. Also, the model is biologically plausible by respecting many biological constraints and important experimental evidence. Finally, the experimental phenomena explained in this paper are all caused by the structure of learned connections between LGN and V1 after the model is trained on natural image data.

2. MATERIALS AND METHODS

2.1. Sparse Coding

The original sparse coding model (Olshausen and Field, 1996) proposed that simple cells represent their sensory input in such a way that their spiking rates in response to natural images tend to be statistically independent and rarely attain large values (near the top of the cells' dynamic range). Mathematically this means that the joint distribution of spike rates over natural images is the product of the distributions for individual cells, and that each of these individual distributions has a long tail (i.e., high kurtosis). Additionally it was proposed that the representation should allow the reconstruction of the sensory input through a simple weighted sum of visual features with minimal error. This can be formulated as an optimization problem of minimizing the cost function,

$$E(\mathbf{A}, \mathbf{s}) = \frac{1}{2}\|\mathbf{x} - \mathbf{A}\mathbf{s}\|_2^2 + \lambda \sum_i Q(s_i), \quad (1)$$

where $\mathbf{x}$ represents the input, columns of the matrix $\mathbf{A}$ represent basis vectors that are universal visual features from which any image can be constructed from a weighted sum, $\mathbf{s}$ is the vector of responses, s_i, of model units that represent the corresponding coefficients for all basis vectors, $Q(\cdot)$ represents a penalty function that favors low activity of model units, and λ is a parameter that scales the penalty function (Olshausen and Field, 1996, 1997). The term $\mathbf{A}\mathbf{s}$ in Equation (1) is the reconstruction of the input from the model, so the first term on the right-hand-side of Equation (1) represents the sum of squared difference between the input and model reconstruction. The second term on the right-hand-side of Equation (1) tends to push $\mathbf{s}$ to small values. Therefore, by solving this minimization problem, the model finds a sparse representation for the input. By taking the partial derivatives of Equation (1) in terms of the elements of $\mathbf{A}$ and $\mathbf{s}$,

and applying gradient descent, the dynamic equations and the learning rule are given by

$$\dot{\mathbf{s}} = \mathbf{A}^T \mathbf{r} - \lambda Q'(\mathbf{s})$$
$$\Delta \mathbf{A} \propto \langle \mathbf{r}\mathbf{s}^T \rangle, \tag{2}$$

where $\mathbf{r} = \mathbf{x} - \mathbf{A}\mathbf{s}$, $\langle \cdot \rangle$ is the average operation, the dot notation represents differentiation with regard to time, and $Q'(\cdot)$ represents the derivative of $Q(\cdot)$.

Based on Equation (2), a network implementation of sparse coding, shown in **Figure 2**, was proposed by Olshausen and Field (1997) who suggested that a feedforward-feedback loop can implement sparse coding. The input to the model was natural images that had been whitened using a filter that resembles the center-surround structure of retinal ganglion RFs. However, the original sparse coding model was not biologically plausible in several aspects, such as the possibility of negative spiking rates and the violation of Dale's law. In addition, the input the the model was not split into separate ON and OFF channels. Finally, this network imposed feedback synaptic connections that were anti-symmetric to the corresponding feedforward connections (i.e., equal but opposite in sign) and it was unclear how such symmetry could be achieved using biologically plausible mechanisms.

2.2. Structure of Our Model

We propose a two-layer network with rate-based neurons that models the activities of LGN cells (first layer), and simple cells (second layer), respectively (**Figure 3**). The model is based on a locally competitive algorithm that efficiently implements sparse coding with neural dynamics with non-negative spiking rates (Rozell et al., 2008).
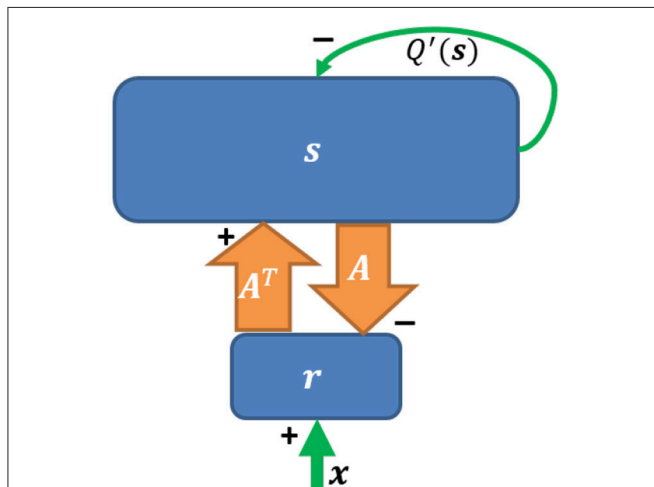
We first define the parameters of the model that will be used throughout the paper. A summary of all symbols defined below is shown in **Table 1**. There are $2N$ LGN cells in the first layer, with N ON LGN cells and N OFF LGN cells, and M simple cells in the second layer. Denote $\mathbf{x} = [x_1, \cdots, x_{2N}]^T$ as the vector of input stimuli to the first layer. Denote $\mathbf{x}_{\text{ON}}$ as the input to ON LGN cells (the first N elements of $\mathbf{x}$) and $\mathbf{x}_{\text{OFF}}$ as the input to OFF LGN cells (the last N elements of $\mathbf{x}$), i.e., $\mathbf{x} = [\mathbf{x}_{\text{ON}}^T, \mathbf{x}_{\text{OFF}}^T]^T$.

Denote $\mathbf{v}^L$ and $\mathbf{s}^L$ as $2N \times 1$ vectors that represent membrane potentials and firing rates of LGN cells in the first layer. Denote $\mathbf{v}_{\text{ON}}^L$, $\mathbf{s}_{\text{ON}}^L$, $\mathbf{v}_{\text{OFF}}^L$, and $\mathbf{s}_{\text{OFF}}^L$ as $N \times 1$ vectors that represent the membrane potentials and firing rates of ON and OFF LGN cells, i.e., $\mathbf{v}^L = [\mathbf{v}_{\text{ON}}^L{}^T, \mathbf{v}_{\text{OFF}}^L{}^T]^T$ and $\mathbf{s}^L = [\mathbf{s}_{\text{ON}}^L{}^T, \mathbf{s}_{\text{OFF}}^L{}^T]^T$. Similarly, $\mathbf{v}^C$ and $\mathbf{s}^C$ are $M \times 1$ vectors that represent membrane potentials and firing rates of M cortical simple cells in the second layer.

In our model, there are several important connections: feedforward (up) excitatory and inhibitory connections from LGN cells to simple cells, feedback (down) excitatory and inhibitory connections from simple cells to LGN cells, and self-excitatory connections of simple cells that represent self-excitation. Definitions of connections are described below. One aspect of the model that lacks biological plausibility is existence of inhibitory connections between thalamus and cortex, but we propose biologically plausible neural circuits of implementing this aspect of the model in the Discussion section.

Denote $\mathbf{A}_{\text{ON}}^{u,+}$ as an $N \times M$ matrix with non-negative elements that represents the feedforward excitatory connections from ON LGN cells to simple cells. Each column of $\mathbf{A}_{\text{ON}}^{u,+}$ represents connections from N ON LGN cells to a simple cell. Similarly, denote $\mathbf{A}_{\text{OFF}}^{u,+}$ as an $N \times M$ matrix with non-negative elements that represents the feedforward excitatory connections from OFF LGN cells to simple cells. Denote $\mathbf{A}_{\text{ON}}^{u,-}$ and $\mathbf{A}_{\text{OFF}}^{u,-}$ as $N \times M$ matrices with non-positive elements that represent inhibitory connections from ON and OFF LGN cells to simple cells, respectively. Denote $\mathbf{A}^{u,+}$ and $\mathbf{A}^{u,-}$ as $2N \times M$ matrices that

FIGURE 2 | The network implementation of sparse coding. Upward and downward arrows represent feedforward and feedback connections. The reconstruction **As** is subtracted via negative feedback. $Q'(\mathbf{s})$ represents self-inhibition of neurons (Adapted from **Figure 5** in Olshausen and Field, 1997).

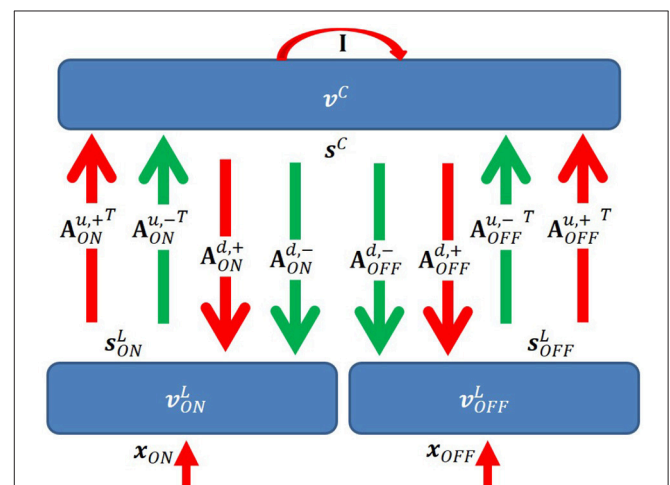**FIGURE 3 |** Graphical representation of the model. **I** is the identity matrix that represents self-excitation. Red and green arrows represent excitatory and inhibitory connections, respectively. Upward and downward arrows are for feedforward and feedback pathways. Notation defined in the main text.

TABLE 1 | Model symbols.

Description	Symbol
Input stimuli to LGN cells	$\mathbf{x}$
Input stimuli to ON LGN cells	$\mathbf{x}_{\text{ON}}$
Input stimuli to OFF LGN cells	$\mathbf{x}_{\text{OFF}}$
Membrane time constant of LGN cells (12 ms)	τ_{L}
Membrane potentials of LGN cells	$\mathbf{v}^{\text{L}}$
Membrane potentials of ON LGN cells	$\mathbf{v}^{\text{L}}_{\text{ON}}$
Membrane potentials of OFF LGN cells	$\mathbf{v}^{\text{L}}_{\text{OFF}}$
Firing rates of LGN cells	$\mathbf{s}^{\text{L}}$
Firing rates of ON LGN cells	$\mathbf{s}^{\text{L}}_{\text{ON}}$
Firing rates of OFF LGN cells	$\mathbf{s}^{\text{L}}_{\text{OFF}}$
Spontaneous firing rate of LGN cells (2 Hz)	s_{b}
Membrane time constant of cortical simple cells (12 ms)	τ_{C}
Membrane potentials of cortical simple cells	$\mathbf{v}^{\text{C}}$
Leakage voltages of cortical simple cells	$\mathbf{v}^{\text{C}}_{\text{leak}}$
Firing rates of cortical simple cells	$\mathbf{s}^{\text{C}}$
Excitatory connection: all LGN cells to simple cells	$\mathbf{A}^{\text{u,+}}$
Excitatory connection: ON LGN cells to simple cells	$\mathbf{A}^{\text{u,+}}_{\text{ON}}$
Excitatory connection: OFF LGN cells to simple cells	$\mathbf{A}^{\text{u,+}}_{\text{OFF}}$
Inhibitory connection: all LGN cells to simple cells	$\mathbf{A}^{\text{u,-}}$
Inhibitory connection: ON LGN cells to simple cells	$\mathbf{A}^{\text{u,-}}_{\text{ON}}$
Inhibitory connection: OFF LGN cells to simple cells	$\mathbf{A}^{\text{u,-}}_{\text{OFF}}$
Excitatory connection: simple cells to all LGN cells	$\mathbf{A}^{\text{d,+}}$
Excitatory connection: simple cells to ON LGN cells	$\mathbf{A}^{\text{d,+}}_{\text{ON}}$
Excitatory connection: simple cells to OFF LGN cells	$\mathbf{A}^{\text{d,+}}_{\text{OFF}}$
Inhibitory connection: simple cells to all LGN cells	$\mathbf{A}^{\text{d,-}}$
Inhibitory connection: simple cells to ON LGN cells	$\mathbf{A}^{\text{d,-}}_{\text{ON}}$
Inhibitory connection: simple cells to OFF LGN cells	$\mathbf{A}^{\text{d,-}}_{\text{OFF}}$
Sparsity level (0.6)	λ
Learning rate	η

represents all excitatory and inhibitory connections from LGN to V1; then we have $\mathbf{A}^{\text{u,+}} = [\mathbf{A}^{\text{u,+}}_{\text{ON}} \ \mathbf{A}^{\text{u,+}}_{\text{OFF}}]$ and $\mathbf{A}^{\text{u,-}} = [\mathbf{A}^{\text{u,-}}_{\text{ON}} \ \mathbf{A}^{\text{u,-}}_{\text{OFF}}]$.

For the feedback pathway, similar notation is used except superscript "d" represents feedback connections from simple cells to LGN cells. Therefore, we have $\mathbf{A}^{\text{d,+}} = [\mathbf{A}^{\text{d,+}}_{\text{ON}} \ \mathbf{A}^{\text{d,+}}_{\text{OFF}}]$ and $\mathbf{A}^{\text{d,-}} = [\mathbf{A}^{\text{d,-}}_{\text{ON}} \ \mathbf{A}^{\text{d,-}}_{\text{OFF}}]$.

Using the notation defined above, the dynamics of ON and OFF LGN cells located in the first layer are given by

$$\tau_{\text{L}}\dot{\mathbf{v}}^{\text{L}}_{\text{ON}} = -\mathbf{v}^{\text{L}}_{\text{ON}} + \mathbf{x}_{\text{ON}} + \mathbf{A}^{\text{d,+}}_{\text{ON}}\mathbf{s}^{\text{C}} + \mathbf{A}^{\text{d,-}}_{\text{ON}}\mathbf{s}^{\text{C}} + s_{\text{b}}$$
$$\mathbf{s}^{\text{L}}_{\text{ON}} = \max(\mathbf{v}^{\text{L}}_{\text{ON}}, 0) \tag{3}$$

and

$$\tau_{\text{L}}\dot{\mathbf{v}}^{\text{L}}_{\text{OFF}} = -\mathbf{v}^{\text{L}}_{\text{OFF}} + \mathbf{x}_{\text{OFF}} + \mathbf{A}^{\text{d,+}}_{\text{OFF}}\mathbf{s}^{\text{C}} + \mathbf{A}^{\text{d,-}}_{\text{OFF}}\mathbf{s}^{\text{C}} + s_{\text{b}},$$
$$\mathbf{s}^{\text{L}}_{\text{OFF}} = \max(\mathbf{v}^{\text{L}}_{\text{OFF}}, 0), \tag{4}$$

where τ_{L} is the time constant of the membrane potentials of LGN cells, s_{b} is a constant that represents the instantaneous firing rate of the background input (i.e., from neurons outside the network),

and the max operation represents the firing dynamics such that a cell only fires when the membrane potential is above a threshold.

Therefore, using the combined notation for ON and OFF LGN cells, the dynamics of LGN cells can be written as

$$\tau_{\text{L}}\dot{\mathbf{v}}^{\text{L}} = -\mathbf{v}^{\text{L}} + \mathbf{x} + (\mathbf{A}^{\text{d,+}} + \mathbf{A}^{\text{d,-}})\mathbf{s}^{\text{C}} + s_{\text{b}}$$
$$\mathbf{s}^{\text{L}} = \max(\mathbf{v}^{\text{L}}, 0). \tag{5}$$

The dynamics of simple cells located in the second layer is given by

$$\tau_{\text{C}}\dot{\mathbf{v}}^{\text{C}} = -(\mathbf{v}^{\text{C}} - \mathbf{v}^{\text{C}}_{\text{leak}}) + \mathbf{A}^{\text{u,+}^T}_{\text{ON}}\mathbf{s}^{\text{L}}_{\text{ON}} + \mathbf{A}^{\text{u,-}^T}_{\text{ON}}\mathbf{s}^{\text{L}}_{\text{ON}}$$
$$+ \mathbf{A}^{\text{u,+}^T}_{\text{OFF}}\mathbf{s}^{\text{L}}_{\text{OFF}} + \mathbf{A}^{\text{u,-}^T}_{\text{OFF}}\mathbf{s}^{\text{L}}_{\text{OFF}} + \mathbf{s}^{\text{C}}, \tag{6}$$

which can be reformulated as

$$\tau_{\text{C}}\dot{\mathbf{v}}^{\text{C}} = -\mathbf{v}^{\text{C}} + \mathbf{v}^{\text{C}}_{\text{leak}} + (\mathbf{A}^{\text{u,+}} + \mathbf{A}^{\text{u,-}})^T\mathbf{s}^{\text{L}} + \mathbf{s}^{\text{C}},$$
$$\mathbf{s}^{\text{C}} = \max(\mathbf{v}^{\text{C}} - \lambda, 0), \tag{7}$$

where τ_{C} is the time constant of the membranes of simple cells and λ is the threshold of the rectifying function of firing rates. In addition, λ is a positive constant that introduces sparseness into the model, $\mathbf{s}^{\text{C}}$ represents the self-excitation of simple cells, which comes from reformulating the model equations of the locally competitive algorithm (Rozell et al., 2008), and $\mathbf{v}^{\text{C}}_{\text{leak}}$ represents the change of membrane potential caused by leakage currents. The leakage currents drive the membrane potentials of simple cells to their resting potentials when there is no external input, i.e., $\mathbf{v}^{\text{C}}$ is zero. Therefore, the steady states of the model dynamics are $\mathbf{v}^{\text{L}} = s_{\text{b}}$, $\mathbf{s}^{\text{L}} = s_{\text{b}}$, $\mathbf{v}^{\text{C}} = 0$, and $\mathbf{s}^{\text{C}} = 0$, which implies that $\mathbf{v}^{\text{C}}_{\text{leak}} = -(\mathbf{A}^{\text{u,+}} + \mathbf{A}^{\text{u,-}})^T\mathbf{s}_{\text{b}}$, where $\mathbf{s}_{\text{b}}$ is a vector whose elements are all equal to s_{b}. Equations 5 and 7 are solved simultaneously by iteration to obtain values of membrane potentials and firing rates.

The codes to run the model are available from ModelDB (http://modeldb.yale.edu/247970).

2.3. Learning Rule

The learning process of the model is based on a Hebbian or anti-Hebbian rule, namely that the change of synaptic strength is related only to local pre-synaptic and post-synaptic activities.

The learning rules are given by

$$\Delta\mathbf{A}^{\text{u,+}} = \eta\langle(\mathbf{s}^{\text{L}} - s_{\text{b}})\mathbf{s}^{\text{C}^T}\rangle$$
$$\Delta\mathbf{A}^{\text{u,-}} = \eta\langle(\mathbf{s}^{\text{L}} - s_{\text{b}})\mathbf{s}^{\text{C}^T}\rangle$$
$$\Delta\mathbf{A}^{\text{d,+}} = -\eta\langle(\mathbf{s}^{\text{L}} - s_{\text{b}})\mathbf{s}^{\text{C}^T}\rangle$$
$$\Delta\mathbf{A}^{\text{d,-}} = -\eta\langle(\mathbf{s}^{\text{L}} - s_{\text{b}})\mathbf{s}^{\text{C}^T}\rangle, \tag{8}$$

where η is the learning rate, $\langle\cdot\rangle$ is the ensemble average operation over some samples, $\mathbf{s}^{\text{L}} - s_{\text{b}}$ is the vector such that each element of vector $\mathbf{s}^{\text{L}}$ is subtracted by scalar s_{b}, and $(\mathbf{s}^{\text{L}} - s_{\text{b}})\mathbf{s}^{\text{C}^T}$ is the matrix given by the outer product of vectors $\mathbf{s}^{\text{L}} - s_{\text{b}}$ and $\mathbf{s}^{\text{C}}$.

The change of synaptic strength depends only on the pre-synaptic activity ($\mathbf{s}^{\text{L}}$) and post-synaptic activity ($\mathbf{s}^{\text{C}}$). Therefore, this learning rule is local and thus biophysically realistic. In

obedience to Dale's law, all the weights of $\mathbf{A}^{u,+}$ and $\mathbf{A}^{d,+}$ are kept non-negative and all weights of $\mathbf{A}^{u,-}$ and $\mathbf{A}^{d,-}$ are kept non-positive during learning. If any synaptic weight changes sign after applying Equation (8), the synaptic weight is set to zero. In addition, after each learning iteration, synaptic weights are multiplicatively normalized to ensure that Hebbian learning is stable. Specifically, each column of $\mathbf{A}^{u,+}$ and $\mathbf{A}^{d,-}$ is normalized to norm l_1 and each column of $\mathbf{A}^{u,-}$ and $\mathbf{A}^{d,+}$ is normalized to norm l_2. The multiplicative normalization of synaptic weights may be achieved by homeostatic mechanisms (Turrigiano, 2011), but these are not implemented here as they are not the focus of this paper.

2.4. Input

The data set used in our simulation consists of 10 pre-whitened 512×512 pixel images of natural scenes provided by Olshausen and Field Olshausen and Field (1996). Some previous studies of sparse coding (efficient coding) also used this data set (Olshausen and Field, 1996; Wiltschut and Hamker, 2009; Zylberberg et al., 2011; Zhu and Rozell, 2013). The input stimuli to the model are chosen to be 16×16 pixel image patches sampled from these 10 pre-whitened 512×512 pixel images, similar to previous studies (Zylberberg et al., 2011; Zhu and Rozell, 2013).

The pre-whitening process mimics the spatial filtering of retinal processing up to a cut-off frequency determined by the limits of visual acuity (Atick and Redlich, 1992). This process is realized by passing the original natural images through a zero-phase whitening filter with root-mean-square power spectrum,

$$R(f) = f e^{-(f/f_c)^4}, \tag{9}$$

where $f_c = 200$ cycles/picture (Olshausen and Field, 1997). **Figure 4** shows the spatial and frequency profiles of the pre-whitening filter. The spatial profile of the filter (**Figure 4C**), obtained by taking the 2D inverse Fourier transform of the filter in the 2D frequency domain, approximates center-surround RFs of LGN cells in a pixel image. The pre-whitening filter described in Equation (9) is widely used in computational studies (Jehee et al., 2006; Jehee and Ballard, 2009; Wiltschut and Hamker, 2009; Zhu and Rozell, 2013).

The pre-whitened images are then scaled to variance 0.2 similar to Olshausen and Field (1997). Image patches are fed into the first layer, which consists of N ON LGN cells and N OFF LGN cells, i.e., one pixel is fed into one ON LGN cell and one OFF LGN cell. If a pixel intensity in a pre-whitened image patch is negative, we assign the absolute value of the pixel intensity to the input of the OFF LGN cell and set the input of the corresponding ON LGN cell to zero; if the pixel intensity is positive, we set the input of the ON LGN cell to the pixel intensity and set the input to the OFF LGN cell to zero.

2.5. Training

Since we use 16×16 pixel images as the input to our model, 256 ON and 256 OFF LGN cells ($N = 256$) are required in the first layer. We use 256 simple cells ($M = 256$) in the second layer. The first-order Euler method is implemented to solve the dynamical system described by Equation 5 and 7. We choose
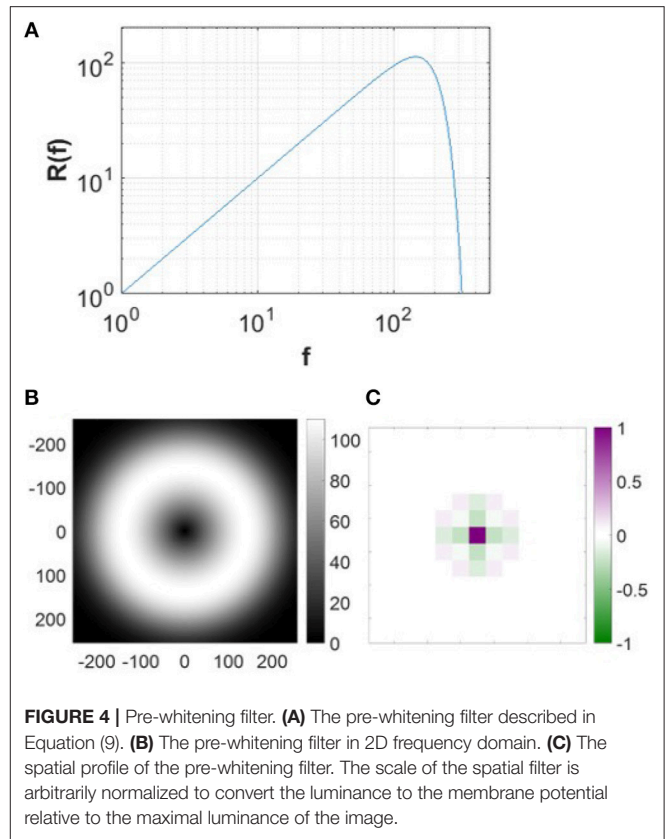

FIGURE 4 | Pre-whitening filter. **(A)** The pre-whitening filter described in Equation (9). **(B)** The pre-whitening filter in 2D frequency domain. **(C)** The spatial profile of the pre-whitening filter. The scale of the spatial filter is arbitrarily normalized to convert the luminance to the membrane potential relative to the maximal luminance of the image.

a time scale in which the passive membrane time constant is $\tau_L = \tau_C = 12$ ms, within the physiological range (Dayan et al., 2001), and sparsity level $\lambda = 0.6$ similar to Zhu and Rozell (2013). The spontaneous firing rate, s_b, is chosen as $s_b = 2$ Hz, the median of spontaneous firing rates of the mouse LGN cells in the experimental study of Tang et al. (2016). There are 30 integration time steps, with an integration time step of 3ms, for calculating neuronal responses per stimulus with the assumption that neural responses will converge after 30 iterations.

Learning rules in Equation (8) are used to update the synaptic weights. For the normalization step after each learning iteration, each column of $\mathbf{A}^{u,+}$ and $\mathbf{A}^{d,-}$ is normalized to have norm l_1 and each column of $\mathbf{A}^{u,-}$ and $\mathbf{A}^{d,+}$ is normalized to have norm l_2. Elements of $\mathbf{A}^{u,+}$ and $\mathbf{A}^{d,+}$ are non-negative and initialized randomly using an exponential distribution with mean 0.5. $\mathbf{A}^{u,-}$ and $\mathbf{A}^{d,-}$ are initialized randomly with non-positive elements that are sampled from an exponential distribution with mean -0.5. Then, synaptic weights are normalized before the learning process starts. Results shown in this paper are from simulations with $l_1 = l_2 = 1$ (unit norm), as used in the previous study by Rozell et al. (2008). The learning rule based on the average activities of a mini-batch is applied; i.e., in every epoch, a mini-batch that consists of 100 randomly selected 16×16 pixel images sampled from the data set is used. Before the training process of natural image patches, the model is pre-trained on white noise for 10, 000 epochs to mimic the process of pre-development of the visual system; the learning rate is 0.5 in pre-training. To ensure that the weights converge after learning on natural

image patches, we use $30,000$ epochs in the training process, where the learning rate is 0.5 for the first $10,000$ epochs, 0.2 for the second $10,000$ epochs and 0.1 for the third $10,000$ epochs. Learning rates were chosen to ensure stable convergence of the weights in a reasonable time; but the results are not sensitive to moderate changes.

2.6. Recovering Receptive Fields of Model Simple Cells Using White Noise

In order to estimate the RFs of model simple cells in a systematic way, we use the method of spike-triggered averaging to find the pattern that each simple cell responds to on average (Schwartz et al., 2006). Using K 16×16 white noise stimuli $\mathbf{n}_1, \cdots, \mathbf{n}_K$, we present pre-processed stimuli to the model, record the firing rates of a simple cell, $s_1, \cdots, s_K$, and then estimate the RF, $\mathbf{F}$, of the simple cell as the weighted average,

$$\mathbf{F} = \frac{s_1 \mathbf{n}_1 + \cdots + s_K \mathbf{n}_K}{s_1 + \cdots + s_K}. \tag{10}$$

We used $70,000$ white noise stimuli, i.e., $K = 70,000$.

In our simulations, we have two versions of estimated RFs using the two different methods of pre-processing the white noise stimuli: the same pre-whitening filter for natural scenes (Equation 9) and a low-pass filter described by

$$L(f) = e^{-(f/f_s)^4}. \tag{11}$$

2.7. Fitting Receptive Fields to Gabor Functions

The RFs of visual cortical cells are often modeled using a 2D Gabor function $G(x, y)$ of the form

$$G(x, y; x_0, y_0, \sigma_x, \sigma_y, f_s, \beta, \theta, \phi)$$
$$= \beta \cos(2\pi f_s x' + \phi) e^{-\left(\frac{x'}{\sqrt{2}\sigma_x}\right)^2 - \left(\frac{y'}{\sqrt{2}\sigma_y}\right)^2} \tag{12}$$

with

$$x' = (x - x_0)\cos\theta + (y - y_0)\sin\theta$$
$$y' = -(x - x_0)\sin\theta + (y - y_0)\cos\theta, \tag{13}$$

where β is the amplitude, (x_0, y_0) is the center, σ_x and σ_y are standard deviations of the Gaussian envelope, θ is the orientation, f_s is the spatial frequency, and ϕ is the phase of the sinusoid wave (Ringach, 2002). These parameters are fitted using the built-in MATLAB (version R2016b, MathWorks, MA, USA) function, *lsqcurvefit*, that efficiently solves non-linear curve-fitting problems using a least-squares method. The fitting error is defined as the square of the ratio between the fitting residual and RF.

To ensure that results were only reported for RFs that were well-fitted to Gabor functions, we excluded RFs for which either (1) the synaptic fields had fitting error larger than 40% or (2) the center of the fitted Gabor functions lay either outside the block, or within one standard deviation of the Gaussian envelope of the block edge (Zylberberg et al., 2011). After applying these two quality control measures, 140 out of 256 model cells remained for subsequent analysis.

2.8. Measuring the Overlap Index Between ON and OFF Sub-regions

To investigate the extent of overlap between ON and OFF sub-regions, we used an overlap index that was used in experimental studies (Schiller et al., 1976; Martinez et al., 2005). Similar to the method used in Martinez et al. (2005), each ON and OFF excitatory sub-region was fitted by an elliptical Gaussian function:

$$h(x, y; x_0, y_0, a, b, \theta, \gamma) = \frac{\gamma}{2\pi ab} e^{-\frac{x'^2}{2a^2} - \frac{y'^2}{2b^2}} \tag{14}$$

where γ is the amplitude, a and b are half axes of the ellipse, and x' and y' are the transformed coordinates given by Equation (13). If there are more than one ON (or OFF) sub-regions for the simple cell, only the most significant sub-region was fitted by the elliptic Gaussian. If either the ON or OFF sub-region of a simple cell has fitting error larger than 40% or has the half axis, a, larger than 3 pixels, this simple cell is excluded. 92 simple cells remained for the analysis of overlap index.

The overlap index, I_o, is then defined as

$$I_o = \frac{W_{ON} + W_{OFF} - d}{W_{ON} + W_{OFF} + d}, \ (-1 < I_o \le 1) \tag{15}$$

where W_{ON} and W_{OFF} are the half width measured at the point where the response is 30% of the maximal response, and d is the distance between the centers of ON and OFF sub-regions. Smaller values of I_o indicate more segregation between ON and OFF sub-regions.

2.9. Measuring the Push-Pull Index

The push-pull effect of the model was measured by a push-pull index (Martinez et al., 2005). First, for each simple cell, we recorded the membrane potential, P, when the preferred input (the synaptic field) was presented to the model. Next, we recorded the membrane potential, N, while presenting the opposite of preferred input to the model. To make the measurement independent of the relative strength of different simple cells, P and N were normalized by

$$P = \frac{P}{\max(|P|, |N|)} \text{ and } N = \frac{N}{\max(|P|, |N|)}. \tag{16}$$

The Push-pull index, I_p, is then defined as

$$I_p = |P + N|, \ (0 \le I_p \le 2). \tag{17}$$

Smaller values of I_p indicate stronger push-pull effect.

2.10. Measuring Contrast Invariance of Orientation Tuning

The method in (Zhu and Rozell, 2013) was used to investigate contrast invariance of orientation tuning and the procedure is as follows. First, an exhaustive search was performed to find the preferred circular sinusoidal grating in the parameter space of the following ranges: radius of the grating was between 1 pixel and $2.5 \min(\sigma_x, \sigma_y)$ (smaller than 8 pixels which is the maximum

radius for a 16 × 16 image patch) with the stepsize of 1 pixel ; spatial frequency was between 0.05 and 0.3 cycles/pixel with the stepsize of 0.05 cycles/pixel; orientation was between 0 and 180 degrees with the stepsize of 5 degrees; phase was between 0 and 360 degrees with the stepsize of 30 degrees. Next, we measured the mean response to the drifting grating with orientations between 0 and 180 degrees with the stepsize of 5 degrees while varying the contrast of the stimuli from 0.2 to 1 in increments of 0.2, where contrast is defined as the amplitude of the sinusoidal grating. The orientation tuning curve for each contrast level was then fit to the Gaussian function and the half-height bandwidth of the Gaussian fit was calculated. The slope of the linear fit to half-height bandwidth vs. contrast for the cell was used to plot the population statistics of contrast invariance (Alitto and Usrey, 2004). Here, only 68 model simple cells that have oriented RFs located well within the 16 × 16 image patch were selected for the analysis.

3. RESULTS

After learning, synaptic weights between LGN and V1 display spatial structures similar to those observed in recordings of neurons in V1, such as oriented Gabor-like filters and non-oriented blobs. Since both excitatory and inhibitory connections from ON and OFF LGN cells contribute to the responses of simple cells, we use the *synaptic field* (S_f) defined as
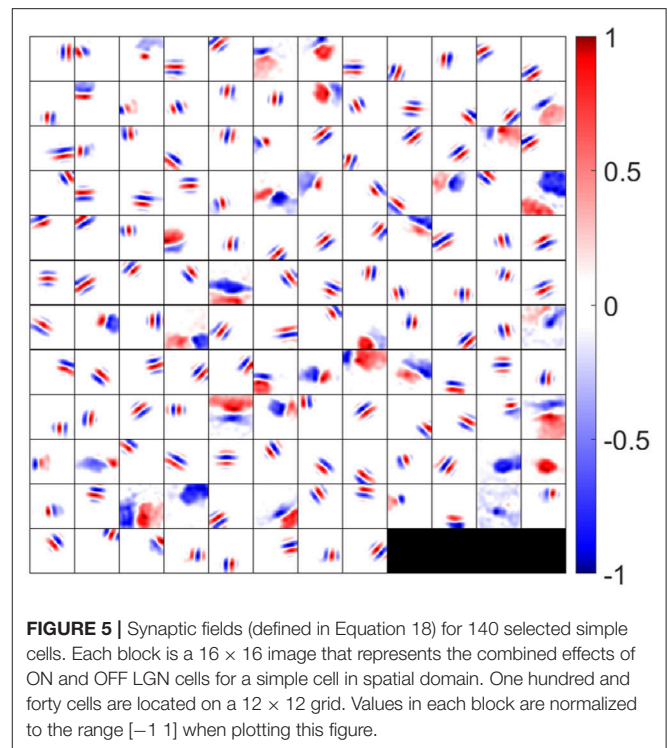
$$S_f = (A_{ON}^{u,+} + A_{ON}^{u,-}) - (A_{OFF}^{u,+} + A_{OFF}^{u,-}) \qquad (18)$$

to visualize the overall synaptic weights from ON and OFF LGN cells. The synaptic fields of 140 model simple cells that meet the two quality control measures (see the Materials and Methods section) are shown in **Figure 5**, where each block represents the overall effect of the feedforward connections from ON and OFF LGN cells to a simple cell. Note that although **Figure 5** displays spatial patterns that are similar to experimental RFs, strictly they represent the synaptic weights from LGN cells to simple cells, which ignores the early visual processing before LGN. However, the RFs of the model are systematically investigated in the following sections.

In the remaining results, we show that the synaptic weights exhibit several properties that have been observed experimentally, including segregation of ON and OFF sub-regions, push-pull effect, phase-reversed feedback, diverse shapes of simple cell RFs, and contrast invariance of orientation tuning.

3.1. Segregated ON and OFF Sub-regions
Hubel and Wiesel found that simple cells in cat striate cortex have spatially separated ON and OFF sub-regions (Hubel and Wiesel, 1962), which was also confirmed by other experimental studies (Jones and Palmer, 1987b; Hirsch et al., 1998; Martinez et al., 2005). However, it is impossible for a model that combines ON and OFF LGN input into a single linear input to explain this important phenomenon. Our model separates ON and OFF LGN input and shows that the learned feedforward excitatory connections from ON and OFF LGN cells to simple cells can

FIGURE 5 | Synaptic fields (defined in Equation 18) for 140 selected simple cells. Each block is a 16 × 16 image that represents the combined effects of ON and OFF LGN cells for a simple cell in spatial domain. One hundred and forty cells are located on a 12 × 12 grid. Values in each block are normalized to the range [−1 1] when plotting this figure.

account for the segregation of ON and OFF sub-regions of simple cells.

ON and OFF excitatory regions of some example simple cells are displayed in **Figure 6A**. In our model, there are 256 ON LGN and 256 OFF LGN cells located evenly on a 16 × 16 image, so each block in **Figure 6A** represents 256 excitatory weights from ON or OFF LGN cells to a simple cell. **Figure 6A** shows that these excitatory connections form spatial patterns such as bars and blobs. Furthermore, a careful examination of the patterns shows that excitatory connections from ON LGN cells are normally adjacent to patterns of excitatory connections from OFF LGN cells, but the two patterns do not overlap, as can be seen when contour plots for the ON and OFF excitatory regions are overlaid in **Figure 6B**.

We quantified the segregation of ON and OFF sub-regions using the overlap index (defined in the Materials and Methods section). The histogram of the overlap index for simple cells in an experimental study (Martinez et al., 2005) is re-plotted in **Figure 6C**. Consistent with the experimental data, 88 out of 92 model simple cells had an overlap index smaller than 0.1 (**Figure 6D**), which indicates that the ON and OFF sub-regions are well-separated in a large majority of the population. The synaptic fields of simple cells whose overlap indices are larger than 0.1 are shown in **Figure 6E**, revealing that most of them have low spatial frequencies.

3.2. Push-Pull Effect
Simple cells are also found to have push-pull responses; i.e., if one contrast polarity excites a cell, the opposite contrast polarity tends to inhibit it (Jones and Palmer, 1987b; Ferster, 1988; Hirsch et al., 1998; Martinez et al., 2005). Even though this effect has been
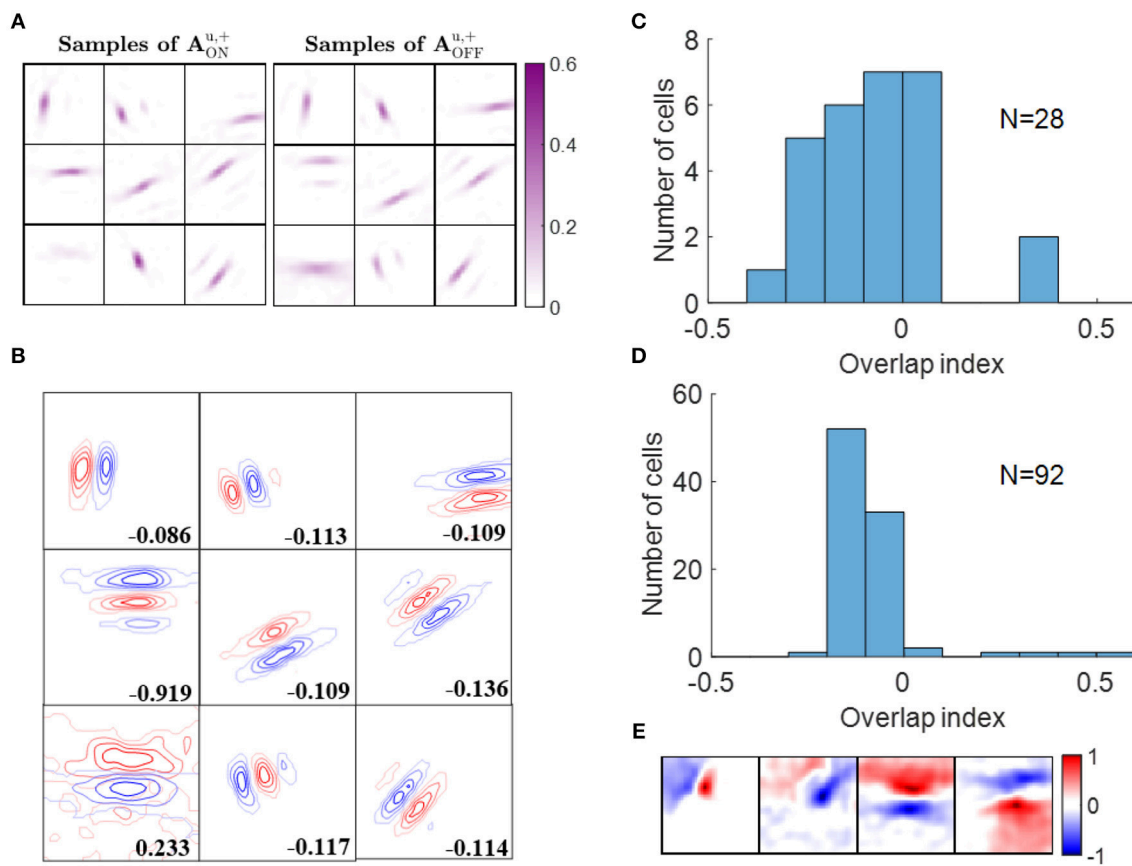
FIGURE 6 | Segregation of ON and OFF sub-regions. **(A)** Some examples of $A_{ON}^{u,+}$ and $A_{OFF}^{u,+}$. Each block is a 16 × 16 image that represents 256 excitatory connections from ON or OFF LGN cells to a simple cell. The color magenta represents excitatory connections. **(B)** Red and blue contours represent excitatory connections from ON and OFF LGN cells, respectively. Connections that are smaller than 20% of the maximal connection were removed to only show the substantial weights. The number in each block indicates the overlap index. **(C)** Histogram of the overlap index for simple cells in cat V1. Re-plotted from Figure 3C in Martinez et al. (2005). **(D)** Histogram of the overlap index for model simple cells. **(E)** Synaptic fields of the four simple cells with overlap index larger than 0.1.

observed in many experimental studies, to our knowledge there has not been a learning model proposed that can explain how this effect emerges. Again, a model that separates ON and OFF LGN input is necessary to investigate the emergence of the push-pull effect. In this section, we show that the push-pull effect for simple cells naturally emerges as a result of neural learning.

Some examples of ON excitatory and OFF inhibitory synaptic weights ($A_{ON}^{u,+}$ and $A_{OFF}^{u,-}$, respectively) are shown in **Figure 7A**. The patterns of $A_{ON}^{u,+}$ are similar to the ones of $A_{OFF}^{u,-}$ and they are located at similar locations, as can be seen from the highly overlapped contours in **Figure 7B**. However, the degree of overlap is different between the examples.

Analogous results to the above also hold for learned excitatory connections from OFF LGN cells, $A_{OFF}^{u,+}$, and inhibitory connections from ON LGN cells, $A_{ON}^{u,-}$ (data not shown).

We then quantified the push-pull effect using push-pull index (defined in the Materials and Methods section). Both the histograms of push-pull index for experimental data (**Figure 7C**) and model simple cells (**Figure 7D**) peaked near zero and showed an decreasing trend. Model simple cells

showed even stronger push-pull index with more simple cells having push-pull index close to zero. The synaptic fields of simple cells with push-pull indices larger than 0.2 are shown in **Figure 7E**, showing that most of them have low spatial frequencies.

3.3. Phase-Reversed Feedback

The experimental study of Wang and colleagues suggests that the synaptic feedback from V1 to LGN is phase-reversed with respect to the feedforward connections (Wang et al., 2006). For example, the connection from a simple cell to an ON-center LGN cell will be excitatory if the ON-center is aligned in visual space to the OFF sub-field of simple cell (i.e., phase-reversed). Conversely, if the ON-center is aligned to the ON sub-field of the simple cell, the connection will be inhibitory. Our learning model with separate ON and OFF LGN cells enables us to investigate the feedback effect from simple cells to LGN cells. In this section, we show that phase-reversed feedback arises in the structures of learned connections.
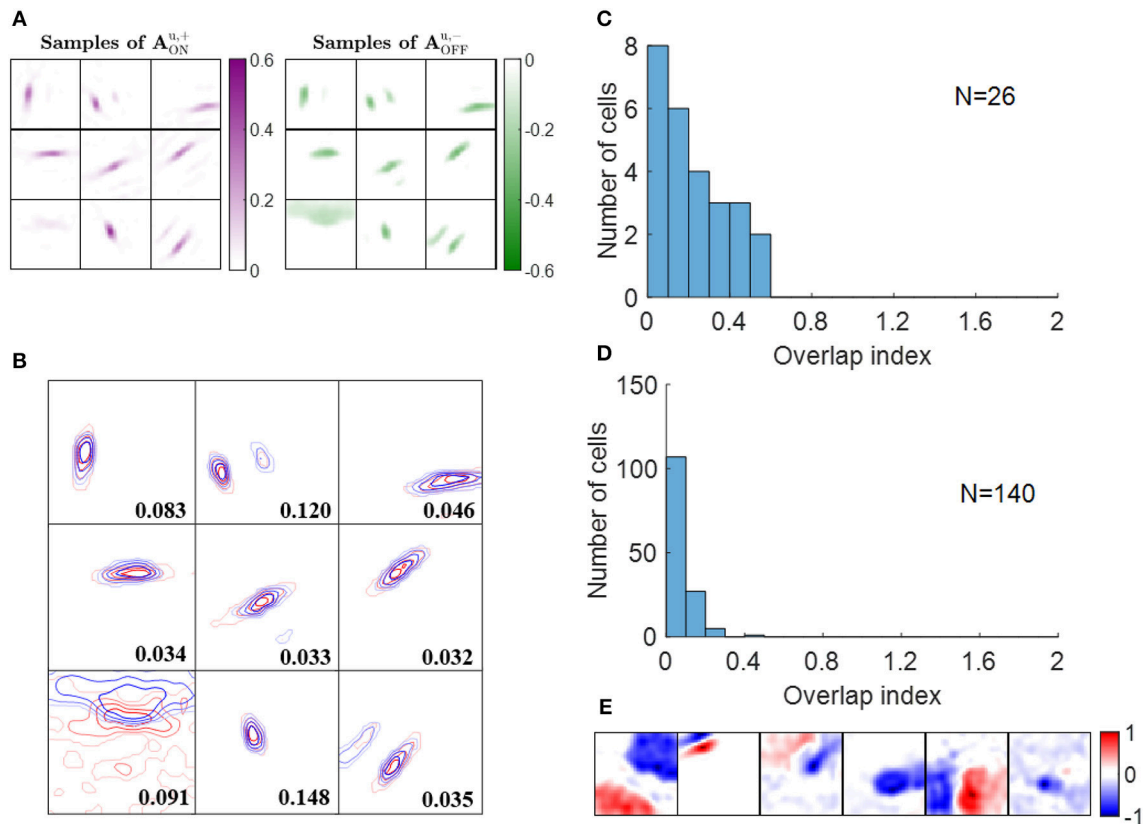
FIGURE 7 | Push-pull effect. (A) Some examples of $\mathbf{A}_{ON}^{u,+}$ and $\mathbf{A}_{OFF}^{u,-}$. Each block on the left is a 16×16 image that represents 256 excitatory connections from ON LGN cells to a simple cell. Each block on the right represents inhibitory connections from OFF LGN cells to a simple cell. The color magenta represents excitatory connections; the color green represents inhibitory connections. **(B)** Red and blue contours represent excitatory connections from ON LGN cells ($\mathbf{A}_{ON}^{u,+}$) and inhibitory connections from OFF LGN cells ($\mathbf{A}_{OFF}^{u,-}$), respectively. Connections that are smaller than 20% of the maximal connection were removed to only show substantial weights. The number in each block indicates the push-pull index. **(C)** Histogram of the push-pull index for simple cells in cat V1. Re-plotted from Figure 4B in Martinez et al. (2005). **(D)** Histogram of the push-pull index for model simple cells. **(E)** Synaptic fields of the six simple cells with push-pull index larger than 0.2.

Feedback from simple cells to LGN cells occurs via both excitatory connections, $\mathbf{A}_x^{d,+}$, and inhibitory connections, $\mathbf{A}_x^{d,-}$, with the overall effect characterized by $\mathbf{A}_x^{d} = \mathbf{A}_x^{d,+} + \mathbf{A}_x^{d,-}$, where x = ON or OFF depending on the type of LGN cell. Therefore, the overall feedback to ON LGN cells, denoted as $\mathbf{A}_{ON}^{d}$, can be represented by $\mathbf{A}_{ON}^{d} = \mathbf{A}_{ON}^{d,+} + \mathbf{A}_{ON}^{d,-}$. Similarly, $\mathbf{A}_{OFF}^{d} = \mathbf{A}_{OFF}^{d,+} + \mathbf{A}_{OFF}^{d,-}$ represents the overall feedback to OFF LGN cells.

The ON and OFF sub-fields of simple cells receptive fields are characterized by the positive and negative regions of the synaptic field defined in Equation (18). The scatter plots in **Figure 8** show that relationship expected from phase-reversed feedback. $\mathbf{S}_f$ is highly positively correlated with $\mathbf{A}_{OFF}^{d}$ (correlation coefficient $r = 0.90$), while $\mathbf{S}_f$ is highly anti-correlated with $\mathbf{A}_{ON}^{d}$ (correlation coefficient $r = -0.92$). According to the figure, wherever $\mathbf{S}_f$ is positive, indicating the ON sub-field, the feedback to ON LGN cells, $\mathbf{A}_{ON}^{d}$, is very likely to be negative and the feedback to OFF LGN cells, $\mathbf{A}_{OFF}^{d}$, tends to be positive; however, wherever $\mathbf{S}_f$ is negative, indicating the OFF-field, the converse is true: the

feedback to ON LGN cells, $\mathbf{A}_{ON}^{d}$, is very likely to be positive and the feedback to OFF LGN cells, $\mathbf{A}_{OFF}^{d}$, tends to be negative. This corresponds to a phase-reversed feedback from V1 to LGN.

This phase-reversed feedback from V1 to LGN can be explained by the learning dynamics of LGN and simple cells described in Equation 8. The learning rule shows that $\mathbf{A}^{u,+}$ and $\mathbf{A}^{d,-}$ are updated with the same magnitude of synaptic change but opposite in sign (and are normalized with the same norm l_1). Similarly, $\mathbf{A}^{u,-}$ and $\mathbf{A}^{d,+}$ are updated with the same magnitude of synaptic change but opposite in sign (and are normalized with the same norm l_2). These anti-symmetries are a consequence of having Hebbian learning for the forward weights and anti-Hebbian learning for the feedback weights. In both cases the magnitude of weight change is proportion to the production of pre- and post-synaptic spike rates, but the sign of the change is opposite. The anti-symmetry arises because roles of pre- and post-synaptic rates are interchanged in forward vs. feedback directions, in combination with the sign change. Simulation results show that $\mathbf{A}^{u,+}$ converges to $-\mathbf{A}^{d,-}$
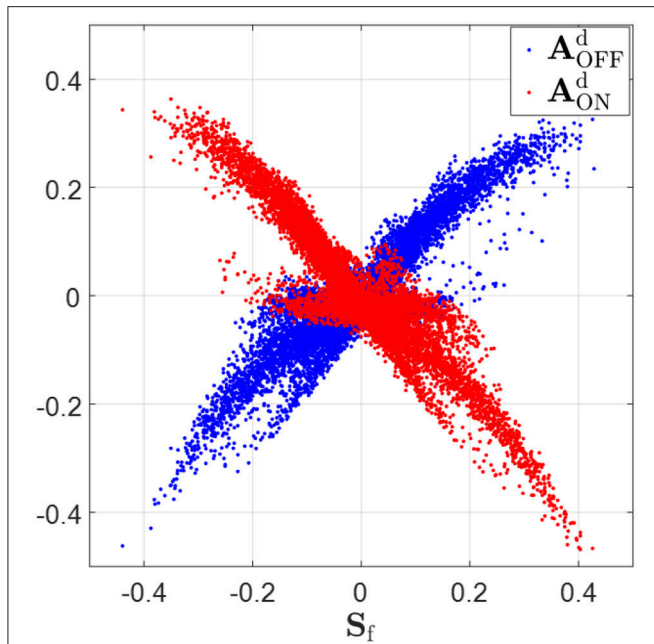
FIGURE 8 | Synaptic fields, $\mathbf{S}_f$ (defined in Equation 18), vs. feedback to ON and OFF LGN cells, $\mathbf{A}_{ON}^d$ and $\mathbf{A}_{OFF}^d$. $\mathbf{S}_f$ is highly positively correlated with $\mathbf{A}_{OFF}^d$ (correlation coefficient $r = 0.90$) and $\mathbf{S}_f$ is highly anti-correlated with $\mathbf{A}_{ON}^d$ (correlation coefficient $r = -0.92$). When $\mathbf{S}_f$ is greater than zero, $\mathbf{A}_{OFF}^d$ tends to be greater than zero and $\mathbf{A}_{ON}^d$ tend to be smaller than zero. On the contrary, $\mathbf{A}_{OFF}^d$ tends to be smaller than zero and $\mathbf{A}_{ON}^d$ tends to be greater than zero if $\mathbf{S}_f$ is negative.
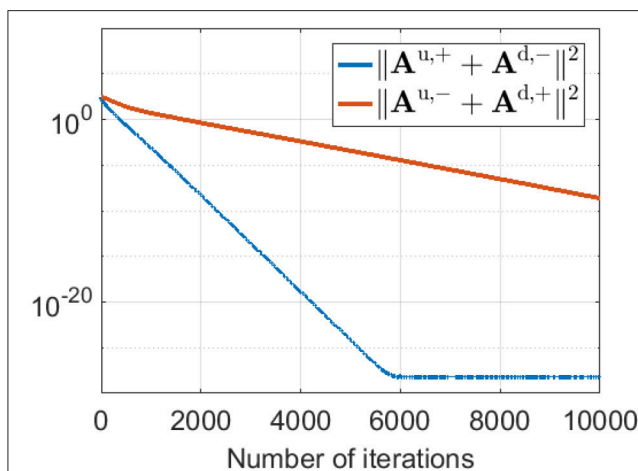

FIGURE 9 | $\|\mathbf{A}^{u,+} + \mathbf{A}^{d,-}\|^2$ and $\|\mathbf{A}^{u,-} + \mathbf{A}^{d,+}\|^2$ during pre-development when white noise is used as the input. The difference between $\mathbf{A}^{u,+}$ and $-\mathbf{A}^{d,-}$ (blue line) decreases to zero very quickly during learning. Similarly, the difference between $\mathbf{A}^{u,-}$ and $-\mathbf{A}^{d,+}$ (red line) reduces to zero quickly, although somewhat slower than the blue line.

and $\mathbf{A}^{u,-}$ converges to $-\mathbf{A}^{d,+}$ even during pre-development when white noise is used as the input to the model, as illustrated in **Figure 9**.

3.4. The Diversity of Model Receptive Fields Resembles That Observed Experimentally for Simple Cells
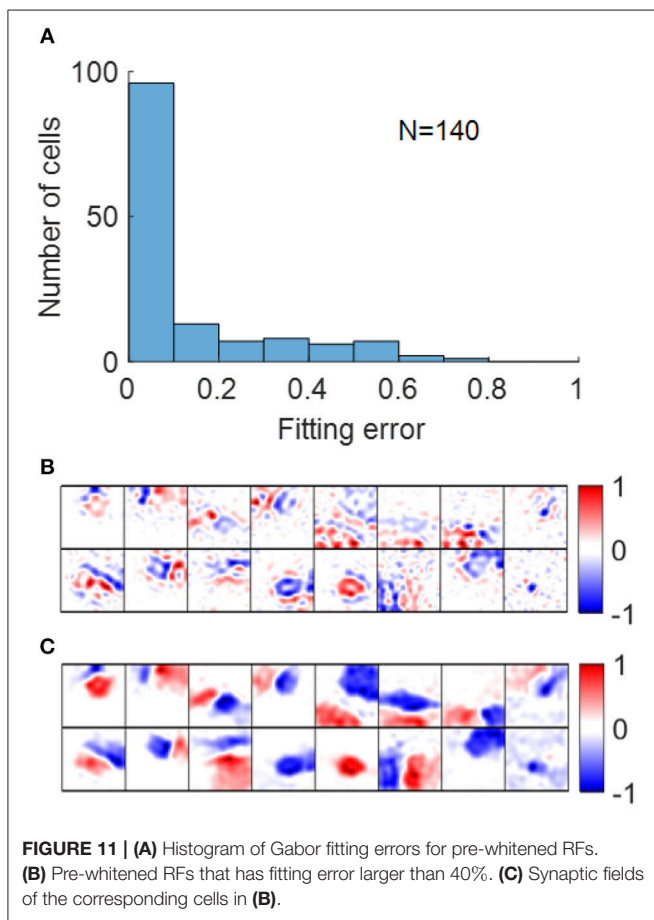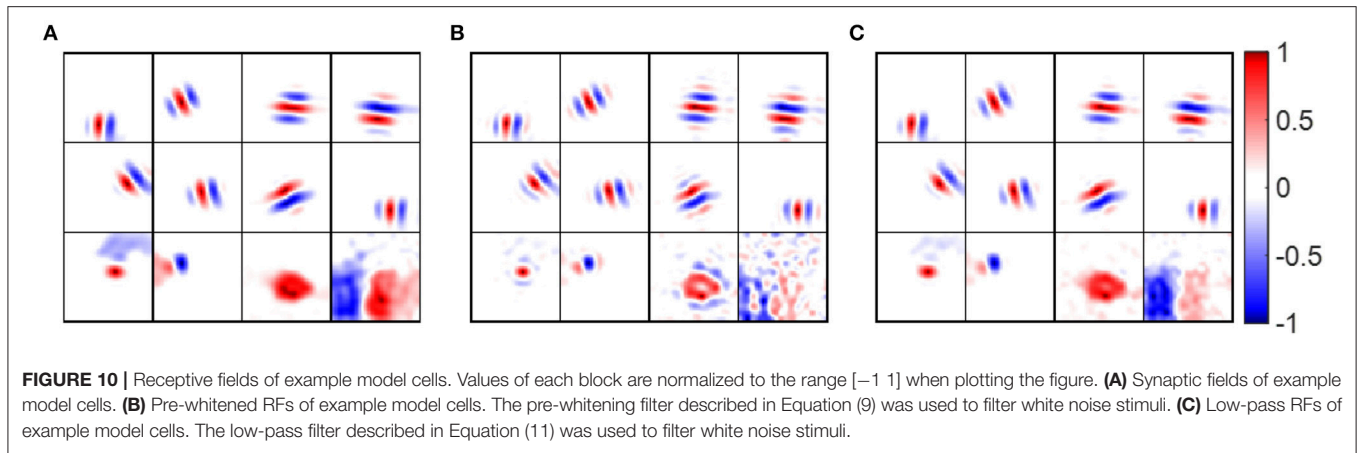
In this section, we show that the range of spatial structures of RFs of our model have a close resemblance to experimental data.

RFs were calculated from the model by simulating experiments in which Gaussian white noise is presented as a visual stimulus, and the spike triggered average is used to estimate RFs. As the presentation of white noise may cause adaptive effects in the early stages visual system relative to natural images, we considered two versions of the model, one with the standard pre-whitening filter (Equation 9) modeling center-surround processing, and a second without pre-whitening in which the filter is replaced by a low-pass filter (Equation 11) with the same upper cut-off frequency as pre-whitening filter. We use *pre-whitened RFs* and *low-pass RFs* to represent of simple cell RFs estimated using the pre-whitening filter and low-pass filter.

Some examples of pre-whitened RFs, low-pass RFs and synaptic fields are shown in **Figure 10**, which shows that pre-whitened RFs and low-pass RFs are similar to synaptic fields. However, pre-whitened RFs tend to have more and thinner stripes, which indicates a narrower tuning to a somewhat higher spatial frequency. For a simple cell tuned to very low spatial frequencies (bottom right blocks), the RF recovered with pre-whitening was a poor match to the original synaptic field, but for RF recovered with low-pass filtering it was fair.

Early studies show that RFs of simple cells can be well-described by 2D Gabor functions described in Equation (12) (Jones and Palmer, 1987a; Ringach, 2002). For our model, most RFs could be well-fitted by Gabor functions with suitable choices of parameters with small fitting errors, as shown in **Figure 11A**. Note that although the fitting error of blob-like RFs might be low, the parameter choices are not necessarily reasonable, in that they are poorly constrained and the process of Gabor fitting imposes an a priori hypothesis that the RF is a 2D-Gabor function even though it is clearly not Gabor-like. The pre-whitened RFs with fitting errors larger than 40% (**Figure 11B**) are cells whose synaptic fields have low spatial frequencies (**Figure 11C**), because pre-whitened RFs of these cells matched poorly to the original synaptic fields (**Figure 10B**). Low-pass RFs of all 140 selected model cells have fitting errors smaller than 40% with 132 of them having fitting errors smaller than 20% (data not shown).

Using fitted parameters of Gabor functions, Ringach constructed a scatter plot of $n_x = \sigma_x f_s$ vs. $n_y = \sigma_y f_s$ to analyze the spatial structures of RFs in V1 over the population (Ringach, 2002). Such plots have subsequently been used by many researchers to investigate the distributions of model simple cell RFs (Rehn and Sommer, 2007; Wiltschut and Hamker, 2009; Zylberberg et al., 2011). n_x and n_y are the width and length of the Gabor function measured in the number of cycles of the spatial frequency (i.e., across and along the stripes). Ringach noted that blob-like RFs are mapped to points near the origin, while RFs with elongated sub-regions are mapped to points away from the origin (Ringach, 2002). In addition, n_x and n_y are directly related with the half-magnitude spatial frequency bandwidth Δf and

FIGURE 10 | Receptive fields of example model cells. Values of each block are normalized to the range [−1 1] when plotting the figure. **(A)** Synaptic fields of example model cells. **(B)** Pre-whitened RFs of example model cells. The pre-whitening filter described in Equation (9) was used to filter white noise stimuli. **(C)** Low-pass RFs of example model cells. The low-pass filter described in Equation (11) was used to filter white noise stimuli.

FIGURE 11 | (A) Histogram of Gabor fitting errors for pre-whitened RFs.
(B) Pre-whitened RFs that has fitting error larger than 40%. **(C)** Synaptic fields of the corresponding cells in **(B)**.

orientation bandwidth $\Delta\theta$ of the fitted Gabor function,

$$
\begin{aligned}
\Delta f := h(n_x) &= \log_2\left(\frac{1 + \frac{\sqrt{2\ln 2}}{2\pi n_x}}{1 - \frac{\sqrt{2\ln 2}}{2\pi n_x}}\right) \text{ in octaves} \\
\Delta\theta := g(n_y) &= 2\arctan\left(\frac{\sqrt{2\ln 2}}{2\pi n_y}\right) \text{ in degrees.}
\end{aligned}
\tag{19}
$$

Both $h(n_x)$ and $g(n_y)$ are monotonically decreasing functions; i.e., the larger n_x and n_y, the smaller Δf and $\Delta\theta$. Note that $h(n_x)$ is not well-defined when $n_x < \sqrt{2\ln 2}/2\pi$ (≈ 0.13), i.e., when the lower half-magnitude frequency do not exist. This corresponds to the region in which Gabor fitting gives ambiguous fits for parameters like spatial frequency and orientation, because oriented RFs with low spatial frequency might lie in this region as well.

We plot n_x vs. n_y and Δf vs. $\Delta\theta$ for RFs obtained from both the model and experimental studies in **Figure 12**. However, the different pre-processing filters for white noise stimuli have a dramatic influence on the distributions of n_x vs. n_y, shifting the distribution for low-pass RFs to the left of pre-whitened RFs, in closer agreement to the experimental data. As mentioned earlier, pre-whitened RFs tend to have more stripes relative to the low-pass RFs, so they are mapped to points away from the origin compared to low-pass RFs. In addition, the distribution of low-pass RFs is continuous from the origin, while there is a gap between points near the origin and points away from the origin for pre-whitened RFs. The inset sub-plots of **Figure 12** show that data points near the origin might be orientated RFs with low spatial frequencies and blob-like RFs might not be necessarily mapped to points near the origin.

In general, oriented RFs are well-described by Gabor functions and low-pass RFs better resemble the distribution of experimental data compared with pre-whitened RFs.

3.5. Contrast Invariance of Orientation Tuning

Another important property of simple cells is contrast invariance of orientation tuning; i.e., the width of the orientation tuning curve is maintained when the contrast of the stimulus changes, as demonstrated in **Figure 13A**. The orientation tuning curves with various stimulus contrasts for a model simple cell are shown in **Figure 13B**, where the bandwidths of each curve remain the same while the responses become larger when the stimulus contrast increases. For a study of contrast invariance of V1 population in ferret, the histogram of the slope of the linear fit of half-width bandwidth vs. contrast (**Figure 13C**) showed that most cells were contrast invariant with the slope close to zero (Alitto and Usrey,
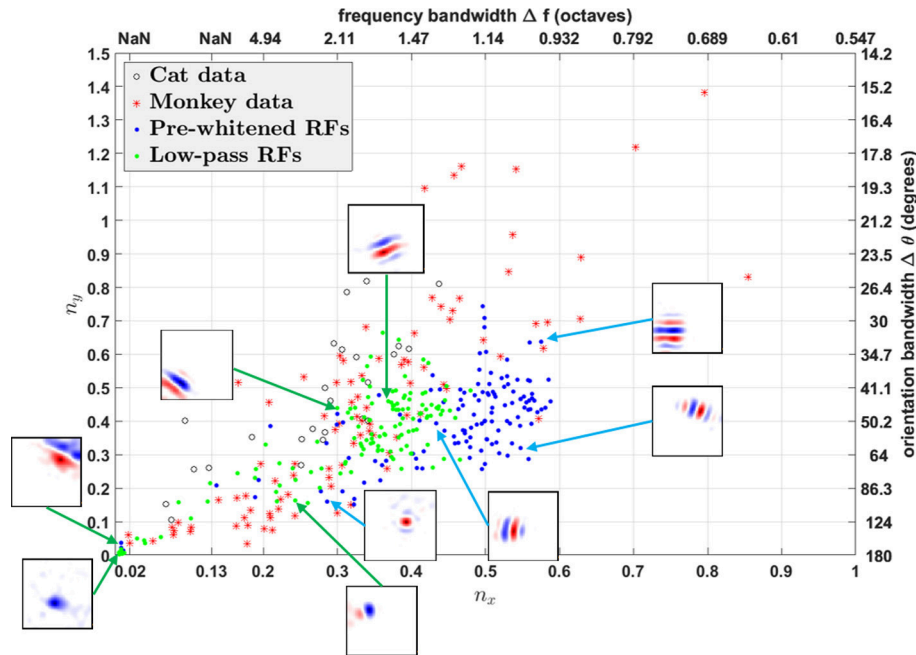
FIGURE 12 | n_x vs. n_y. Comparison of RFs of the model with experimentally recorded data for cat simple cells and monkey simple cells. Open circles: 25 cat simple cells from Table 1 in Jones and Palmer (1987a) re-plotted in the (n_x, n_y) plane; red stars: 93 monkey simple cells in Ringach (2002); blue dots: pre-whitened RFs using the pre-whitening filter described in Equation (9); green dots: low-pass RFs using the low-pass filter described in Equation (11). The axes on the top and right represent frequency and orientation bandwidths of fitted Gabor functions computed using Equation (19). Some examples of RFs are displayed in the inset sub-plots. Data points of estimated RFs with fitting errors > 40% were excluded, which gave 124 data points for pre-whitened RFs and 140 data points for low-pass RFs.

2004). **Figure 13D** shows that most model cells have the slope around zero, which is consistent with experimental data.

4. DISCUSSION

4.1. Relationship With Sparse Coding

Sparse coding has been successful in modeling simple cell receptive fields (RFs) and has been used by many researchers over the past years. Our model is based on an algorithm that efficiently implements sparse coding (Rozell et al., 2008), and is therefore closely related to the original concept of sparse coding (Olshausen and Field, 1996).

If we define $\mathbf{A}$ as a $2N \times M$ matrix that represents the overall effect caused by excitatory and inhibitory connections from $2N$ LGN cells to M simple cells, we have $\mathbf{A} = \mathbf{A}^{u,+} + \mathbf{A}^{u,-}$. The dynamics of simple cells described in Equation (7) can be rewritten as

$$\tau_C \dot{\mathbf{v}}^C = -\mathbf{v}^C + \mathbf{A}^T(\mathbf{s}^L - s_b) + \mathbf{s}^C. \tag{20}$$

As illustrated in **Figure 9**, $\mathbf{A}^{u,+} \rightarrow -\mathbf{A}^{d,-}$ and $\mathbf{A}^{u,-} \rightarrow -\mathbf{A}^{d,+}$ during learning. Therefore, we have $\mathbf{A}^{d,-} + \mathbf{A}^{d,+} = -\mathbf{A}^{u,+} - \mathbf{A}^{u,-} = -\mathbf{A}$. The dynamics of LGN cells described in Equation (5) can be rewritten as

$$\tau_L \dot{\mathbf{v}}^L = -\mathbf{v}^L + \mathbf{x} - \mathbf{A}\mathbf{s}^C + s_b. \tag{21}$$

If the columns of $\mathbf{A}$ are seen as the basis vectors of a generative model, $\mathbf{A}\mathbf{s}^C$ can be seen as the linear reconstruction of the

input using learned basis vectors and thus $\mathbf{x} - \mathbf{A}\mathbf{s}^C$ represents the residual error, which is similar to $\mathbf{r}$ of the sparse coding formulation given in Equation (2). Therefore, the residual error used to update the basis vectors of the original sparse coding model is represented by the responses of LGN cells in our model.

To incorporate Dale's law, non-negative connections, $\mathbf{A}^{u,+}$, and non-positive connections, $\mathbf{A}^{u,-}$, are employed in our model to represent the positive and negative elements of $\mathbf{A}$. $\mathbf{A}^{u,+}$ and $\mathbf{A}^{u,-}$ are not co-active in general, which suggests that $\mathbf{A}^{u,+} \approx [\mathbf{A}]_+$ and $\mathbf{A}^{u,-} \approx [\mathbf{A}]_-$, where $[\cdot]_+$ preserves the positive elements and sets negative elements to zero and $[\cdot]_-$ preserves the negative elements and sets positive elements to zero.

In other words, our model is essentially a variant of sparse coding that employs separate connections to learn the positive and negative part of the overall connections.

4.2. Relationship With Predictive Coding

Our model is a hierarchical model with feedforward and feedback connections based on a locally competitive algorithm (Rozell et al., 2008). The structure of our model is essentially very similar to that of predictive coding models. To be more specific, the feedback from the second-layer neurons reconstruct the input. The residual error is computed at the first layer and then propagated to the second layer via feedforward connections.

Although our model presented here and the predictive coding model of Jehee and Ballard (2009) can explain phase-reversed feedback, the models differ in several respects. First, sparse
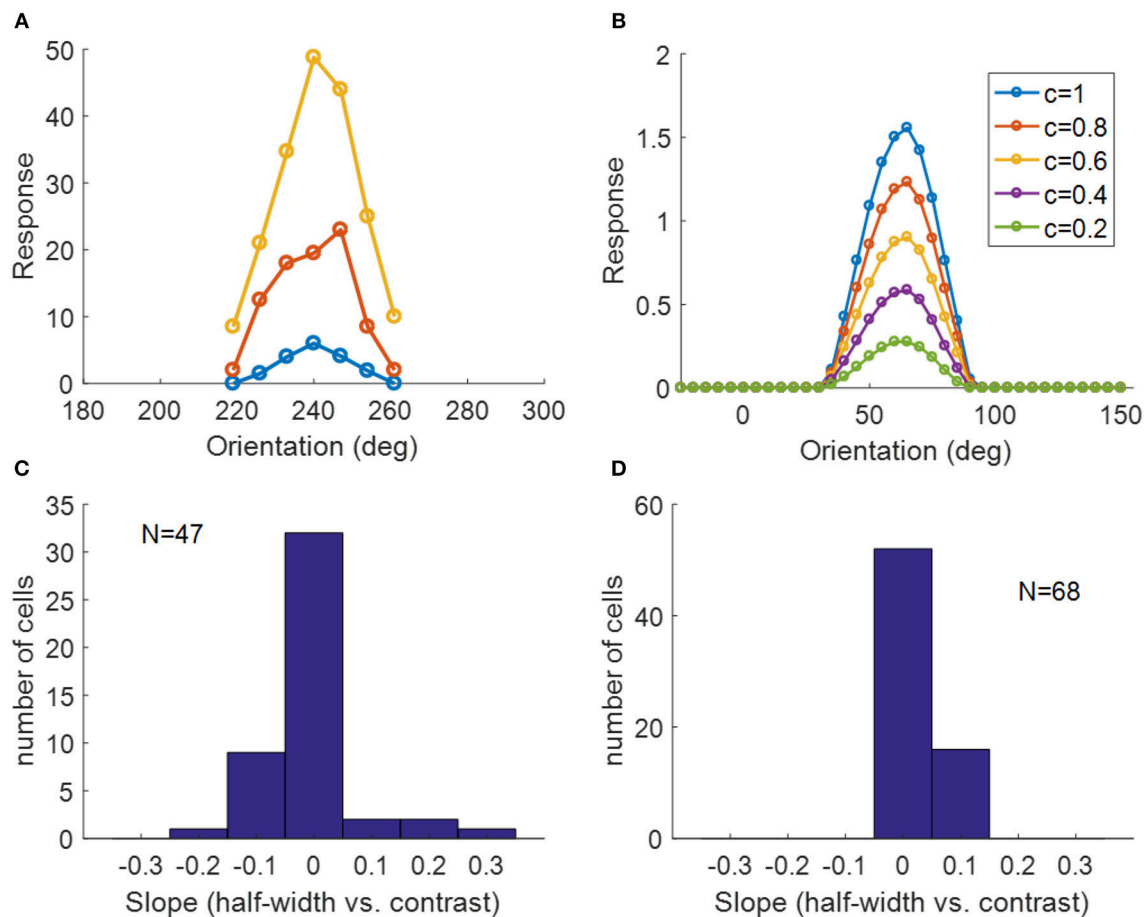
FIGURE 13 | Contrast invariance of orientation tuning. **(A)** Contrast invariant orientation tuning curves of a simple cell in cat V1. Re-plotted from Figure 3A in Skottun et al. (1987). Different colors represent different contrasts. **(B)** Contrast invariant orientation tuning curves of a cell in our model. $c = 1$ and $c = 0.2$ correspond to the high and low contrast, respectively. **(C)** Histogram of the slope of half-height bandwidth vs. contrast for V1 population in ferret. Re-plotted from Figure 3B in Alitto and Usrey (2004). **(D)** Histogram of the slope of half-height bandwidth vs. contrast for model simple cells.

coding in our model is simply realized by the threshold of the rectifying function of firing rates for simple cells and this simple mechanism leads to simple neural circuits. Second, compared to the mechanism for determining simple cell responses one by one in their model, our model computes the responses in parallel. Third, our model generates diverse types of RFs that correspond well to experimental data. Finally, the phase-reversed effect is simply accounted for by the special pattern of learned connections, which also explains the segregation of ON/OFF sub-regions and push-pull effect for simple cells.

4.3. The Function of Spontaneous Activity

In the model proposed here, the dynamics of LGN cells described in Equation (5) has the background firing rate, s_b, as part of the input to LGN cells. This spontaneous firing rate introduces a shift of the operating point for LGN cells. Given the responses of simple cells, s^C, $x - As^C$ in Equation (21) represents the reconstruction residual error between the input and reconstruction. The residual error gives the difference between the real input and the representation produced by the

model and it can be either positive or negative. To code for the signed quantities (residual error), Ballard and Jehee carried out a case-by-case study, leading to very complicated neural circuits (Ballard and Jehee, 2012). However, our model has a straightforward method for the implementation of solving signed quantities. The background firing rate, s_b, in Equation (5) increases the residual errors by s_b. Therefore, the membrane potential of LGN cell, v^L, represents the residual error shifted up by s_b. The threshold function in Equation (5) gives the firing rate of the LGN cell and it preserves the residual error in the interval of $[-s_b, \infty]$, which preserves the information of whether the model under-estimates or over-estimates the input stimuli and forces the connections to evolve through learning in the correct direction. In Equation (7), which describes simple cell dynamics, the effect of the spontaneous firing rate, s_b, is removed by v^C_{leak}, a homeostatic mechanism employed by simple cells to maintain resting membrane potentials when there is no external input. The local learning rule described by Equation (8) also eliminates the effect of the spontaneous firing rate by subtracting it. The use of spontaneous firing rate makes the model much simpler

and offers a new approach for solving the problem of signed quantities (residual errors). Experimental evidence shows that thalamocortical neurons can fire with bursts of action potentials without any synaptic input (Kandel et al., 2013), which suggests that the spontaneous firing activities might be used to encode the difference between input and feedback information.

4.4. Pre-processing of the Early Visual System

Atick and Redlich suggest that the retinal goal is to whiten the visual input up to a transition frequency such that input noise can also be suppressed (Atick and Redlich, 1992). The pre-whitening filter (Equation 9) approximately whitens the natural scenes up to the cut-off frequency.

However, for pre-processing white noise stimuli, two hypotheses are considered here. First, the filtering process of the early visual system can be described by the pre-whitening filter (Equation 9) whether or not the visual stimuli are natural scenes. Second, the early visual system is adaptive such that the visual stimuli are whitened up to a cut-off frequency. In this case, a low-pass filter (Equation 9) should be used, because white noise stimuli are already whitened across all frequencies. Our results suggest that estimated RFs using low-passed white noise match the experimental data much better than estimated RFs using pre-whitened white noise. Further investigation of how visual stimuli are processed before they are fed to the visual cortex is needed to better understand the properties of simple cells.

4.5. The Role of l_1 and l_2

Each column of $\mathbf{A}^{u,+}$ and $\mathbf{A}^{d,-}$ is normalized to norm l_1 and each column of $\mathbf{A}^{u,-}$ and $\mathbf{A}^{d,+}$ is normalized to norm l_2. In other words, l_1 represents the overall strength of feedforward excitatory connections and feedback inhibitory connections while l_2 represents the overall strength of feedforward inhibitory connections and feedback excitatory connections. The results shown in this paper are based on $l_1 = 1$ and $l_2 = 1$; i.e., the strength of feedforward excitatory connections is equivalent to feedforward inhibitory connections, which leads to a strong push-pull effect in **Figure 7D**. If l_2 is smaller than l_1, the push-pull effect will be weaker and the distribution of the push-pull index will shift to the right. In addition, reducing l_2 results in more blob-like receptive fields (data not shown).

4.6. Neural Circuits

Biologically realistic neural models can provide deeper insights into how real neural circuits function. The model proposed here contains a number of features that correspond to those in its biological counterpart, namely in terms of ON and OFF channels for LGN cells, positive neuronal responses, local computation, local learning rule, existence of feedback, and obedience to Dale's law.

In addition, our model incorporates inhibitory effects between LGN cells and cortical simple cells. As pointed out in the Materials and Methods section, for simplicity, inhibitory effects are implemented by direct inhibitory connections between two layers. However, in reality, long-range inhibitory effects should be implemented via interneurons that have

inhibitory synapses. In this section, we will discuss several neural circuits of implementing inhibitory connections of our model.

Possible neural circuits that may be used to implement long-range inhibition are displayed in **Figure 14**. Assume that the overall inhibitory effects from LGN cells (with activity $\mathbf{s}^L$) to cortical simple cells (with activity $\mathbf{s}^C$) can be represented by inhibitory connections, $\mathbf{A}^-$, between populations. We also assume that the learning rule of $\mathbf{A}^-$ is local, i.e., that only depends on the responses of two populations ($\mathbf{s}^L$ and $\mathbf{s}^C$). Long-range inhibition in our model is implemented via direct inhibitory connections, which is not biologically realistic (**Figure 14A**).

The circuit in **Figure 14B** implements inhibitory connections, $\mathbf{A}^-$ (with non-positive weights), via a population of interneurons that have inhibitory connections, $\mathbf{A}^-$, with cortical simple cells. LGN cells are connected to interneurons via long-range identical excitatory connections, $\mathbf{I}$; i.e., the interneurons copy the responses of LGN cells. For this structure, long-range excitatory connections, $\mathbf{I}$, are fixed while $\mathbf{A}^-$ are learned using the same learning rule in **Figure 14A**. In this case, the learning rule of $\mathbf{A}^-$ is still local because the responses of interneurons are just $\mathbf{s}^L$ and the model is still biologically plausible in terms of the local learning rule. Furthermore, the RFs of interneurons in the same layer as cortical simple cells should be LGN-like. Though V1 cortical cells with blob-like RFs were found in different species (Kretz et al., 1986; Jones and Palmer, 1987a; Hawken et al., 1988; Muly and Fitzpatrick, 1992; Chapman and Stryker, 1993;
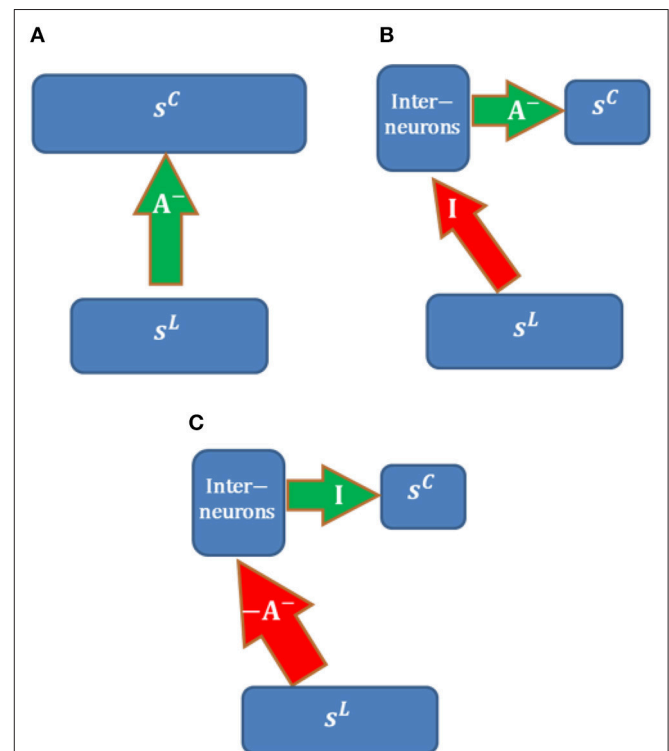

FIGURE 14 | Possible neural circuits for implementing long-range inhibition. Red and green arrows represent excitatory and inhibitory connections. **(A)** Direct long-range inhibition. **(B)** Circuit I. **(C)** Circuit II.

Ringach, 2002), we are not sure whether this neural circuit is the most likely candidate because the fixed identical connection between LGN cells and the interneurons seems artificial unless they can be learned.

Figure 14C shows another possible neural circuit for implementing $\mathbf{A}^-$. LGN Cells are connected to interneurons via long-range excitatory connections, $-\mathbf{A}^-$. There is a one-to-one mapping between interneurons and cortical simple cells. In this case, the overall effect from LGN cells to simple cells is equivalent to $\mathbf{A}^-$. In addition, the RFs of inhibitory interneurons should resemble simple cells and show orientation tuning since the learned $\mathbf{A}^-$ has spatial structures such as oriented bars, which is consistent with the smooth simple cells found in cat V1 of the experimental study (Hirsch et al., 2003). The positive connections $-\mathbf{A}^-$ can be learned by Hebbian learning and the identical connections between interneurons and cortical simple cells can be learned by anti-Hebbian learning. Therefore, this neural circuit is more feasible than than the circuit in **Figure 14B**.

4.7. Discrepancies Between Model and Experimental Data

Our model can capture the most significant features of experimental phenomena such as the segregation of ON and OFF sub-regions, push-pull effect and contrast invariance of orientation tuning. However, there are also discrepancies between the distributions of model and experimental data. In general, the histograms of experimental data (**Figures 6C, 7C, 13C**) are wider than model data (**Figures 6D**, **7D**, **13D**), which shows that experimental data is more diverse. One possible explanation is that model cells in this paper are only a subset of the rich repository of real cortical cells. Furthermore, choices of free parameters in the model might also lead to different results.

5. CONCLUSION

In this paper, we presented a biologically plausible model of LGN-V1 pathways to account for many experimental phenomena of V1. We found that the segregation of ON/OFF sub-regions of simple cells, push-pull effect, and phase-reversed cortico-thalamic feedback can all be explained by the structure of learning connections when the model incorporates ON and OFF LGN cells and is trained using natural images. Furthermore, the model can produce diverse shapes of receptive fields and contrast invariance of orientation tuning of simple cells, consistent with experimental observations.

DATA AVAILABILITY

The datasets generated for this study are available on request to the corresponding author.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

REFERENCES

Alitto, H. J., and Usrey, W. M. (2004). Influence of contrast on orientation and temporal frequency tuning in ferret primary visual cortex. *J. Neurophysiol.* 91, 2797–2808. doi: 10.1152/jn.00943.2003

Atick, J. J., and Redlich, A. N. (1992). What does the retina know about natural scenes? *Neural Comput.* 4, 196–210. doi: 10.1162/neco.1992.4.2.196

Ballard, D. H., and Jehee, J. (2012). Dynamic coding of signed quantities in cortical feedback circuits. *Front. Psychol.* 3:254. doi: 10.3389/fpsyg.2012.00254

Chapman, B., and Stryker, M. P. (1993). Development of orientation selectivity in ferret visual cortex and effects of deprivation. *J. Neurosci.* 13, 5251–5262. doi: 10.1523/JNEUROSCI.13-12-05251.1993

Dayan, P., Abbott, L. F., and Abbott, L. (2001). *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*. Cambridge, MA: MIT Press.

Ferster, D. (1988). Spatially opponent excitation and inhibition in simple cells of the cat visual cortex. *J. Neurosci.* 8, 1172–1180. doi: 10.1523/JNEUROSCI.08-04-01172.1988

Ferster, D., Chung, S., and Wheat, H. (1996). Orientation selectivity of thalamic input to simple cells of cat visual cortex. *Nature* 380, 249–252. doi: 10.1038/380249a0

Finn, I. M., Priebe, N. J., and Ferster, D. (2007). The emergence of contrast-invariant orientation tuning in simple cells of cat visual cortex. *Neuron* 54, 137–152. doi: 10.1016/j.neuron.2007.02.029

Hawken, M. J., Parker, A. J., and Lund, J. (1988). Laminar organization and contrast sensitivity of direction-selective cells in the striate

cortex of the old world monkey. *J. Neurosci.* 8, 3541–3548. doi: 10.1523/JNEUROSCI.08-10-03541.1988

Hirsch, J. A., Alonso, J. M., Reid, R. C., and Martinez, L. M. (1998). Synaptic integration in striate cortical simple cells. *J. Neurosci.* 18, 9517–9528. doi: 10.1523/JNEUROSCI.18-22-09517.1998

Hirsch, J. A., Martinez, L. M., Pillai, C., Alonso, J. M., Wang, Q., and Sommer, F. T. (2003). Functionally distinct inhibitory neurons at the first stage of visual cortical processing. *Nat. Neurosci.* 6:1300. doi: 10.1038/nn1152

Hubel, D. H., and Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *J. Physiol.* 148, 574–591. doi: 10.1113/jphysiol.1959.sp006308

Hubel, D. H., and Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol.* 160, 106–154. doi: 10.1113/jphysiol.1962.sp006837

Hubel, D. H., and Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *J. Physiol.* 195, 215–243. doi: 10.1113/jphysiol.1968.sp008455

Jehee, J. F., and Ballard, D. H. (2009). Predictive feedback can account for biphasic responses in the lateral geniculate nucleus. *PLoS Comput. Biol.* 5:e1000373. doi: 10.1371/journal.pcbi.1000373

Jehee, J. F., Rothkopf, C., Beck, J. M., and Ballard, D. H. (2006). Learning receptive fields using predictive feedback. *J. Physiol.* 100, 125–132. doi: 10.1016/j.jphysparis.2006.09.011

Jin, J., Wang, Y., Swadlow, H. A., and Alonso, J. M. (2011). Population receptive fields of on and off thalamic inputs to an orientation column in visual cortex. *Nat. Neurosci.* 14, 232–240. doi: 10.1038/nn.2729

Jin, J. Z., Weng, C., Yeh, C. I., Gordon, J. A., Ruthazer, E. S., Stryker, M. P., et al. (2008). On and off domains of geniculate afferents in cat primary visual cortex. *Nat. Neurosci.* 11, 88–94. doi: 10.1038/nn2029

Jones, J. P., and Palmer, L. A. (1987a). An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. *J. Neurophysiol.* 58, 1233–1258. doi: 10.1152/jn.1987.58.6.1233

Jones, J. P., and Palmer, L. A. (1987b). The two-dimensional spatial structure of simple receptive fields in cat striate cortex. *J. Neurophysiol.* 58, 1187–1211. doi: 10.1152/jn.1987.58.6.1187

Kandel, E., Schwartz, J., Jessell, T., Siegelbaum, S., and Hudspeth, A. (2013). *Principles of Neural Science, 5th Edn.* New York, NY: McGraw-Hill Education.

Kretz, R., Rager, G., and Norton, T. (1986). Laminar organization of ON and OFF regions and ocular dominance in the striate cortex of the tree shrew (*Tupaia belangeri*). *J. Comp. Neurol.* 251, 135–145. doi: 10.1002/cne.902510110

Lee, W. C. A., Bonin, V., Reed, M., Graham, B. J., Hood, G., Glattfelder, K., et al. (2016). Anatomy and function of an excitatory network in the visual cortex. *Nature* 532, 370–374. doi: 10.1038/nature17192

Martinez, L. M., Wang, Q., Reid, R. C., Pillai, C., Alonso, J. M., Sommer, F. T., et al. (2005). Receptive field structure varies with layer in the primary visual cortex. *Nat. Neurosci.* 8, 372–379. doi: 10.1038/nn1404

Muly, E. C., and Fitzpatrick, D. (1992). The morphological basis for binocular and ON/OFF convergence in tree shrew striate cortex. *J. Neurosci.* 12, 1319–1334. doi: 10.1523/JNEUROSCI.12-04-01319.1992

Olshausen, B. A., Cadieu, C. F., and Warland, D. K. (2009). "Learning real and complex overcomplete representations from the statistics of natural images," in *Proceedings of SPIE - The International Society for Optical Engineering* (San Diego, CA).

Olshausen, B. A., and Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* 381, 607–609. doi: 10.1038/381607a0

Olshausen, B. A., and Field, D. J. (1997). Sparse coding with an overcomplete basis set: a strategy employed by V1? *Vision Res.* 37, 3311–3325. doi: 10.1016/S0042-6989(97)00169-7

Priebe, N. J. (2016). Mechanisms of orientation selectivity in the primary visual cortex. *Annu. Rev. Vis. Sci.* 2, 85–107. doi: 10.1146/annurev-vision-111815-114456

Rehn, M., and Sommer, F. T. (2007). A network that uses few active neurones to code visual input predicts the diverse shapes of cortical receptive fields. *J. Comput. Neurosci.* 22, 135–146. doi: 10.1007/s10827-006-0003-9

Ringach, D. L. (2002). Spatial structure and symmetry of simple-cell receptive fields in macaque primary visual cortex. *J. Neurophysiol.* 88, 455–463. doi: 10.1152/jn.2002.88.1.455

Rozell, C. J., Johnson, D. H., Baraniuk, R. G., and Olshausen, B. A. (2008). Sparse coding via thresholding and local competition in neural circuits. *Neural Comput.* 20, 2526–2563. doi: 10.1162/neco.2008.03-07-486

Schiller, P. H., Finlay, B. L., and Volman, S. F. (1976). Quantitative studies of single-cell properties in monkey striate cortex. I. spatiotemporal organization of receptive fields. *J. Neurophysiol.* 39, 1288–1319. doi: 10.1152/jn.1976.39.6.1288

Schwartz, O., Pillow, J. W., Rust, N. C., and Simoncelli, E. P. (2006). Spike-triggered neural characterization. *J. Vision* 6, 484–507. doi: 10.1167/6.4.13

Sclar, G., and Freeman, R. (1982). Orientation selectivity in the cat's striate cortex is invariant with stimulus contrast. *Exp. Brain Res.* 46, 457–461. doi: 10.1007/BF00238641

Sherman, S. M., and Guillery, R. (1996). Functional organization of thalamocortical relays. *J. Neurophysiol.* 76, 1367–1395. doi: 10.1152/jn.1996.76.3.1367

Skottun, B. C., Bradley, A., Sclar, G., Ohzawa, I., and Freeman, R. D. (1987). The effects of contrast on visual orientation and spatial frequency discrimination: a comparison of single cells and behavior. *J. Neurophysiol.* 57, 773–786. doi: 10.1152/jn.1987.57.3.773

Strata, P., and Harvey, R. (1999). Dale's principle. *Brain Res. Bull.* 50, 349–350. doi: 10.1016/S0361-9230(99)00100-8

Swadlow, H. A. (1983). Efferent systems of primary visual cortex: a review of structure and function. *Brain Res. Rev.* 6, 1–24. doi: 10.1016/0165-0173(83)90002-4

Tang, J., Ardila Jimenez, S. C., Chakraborty, S., and Schultz, S. R. (2016). Visual receptive field properties of neurons in the mouse lateral geniculate nucleus. *PLoS ONE* 11:e0146017. doi: 10.1371/journal.pone.0146017

Turrigiano, G. (2011). Too many cooks? intrinsic and synaptic homeostatic mechanisms in cortical circuit refinement. *Ann. Rev. Neurosci.* 34, 89–103. doi: 10.1146/annurev-neuro-153238

Wang, W., Jones, H. E., Andolina, I. M., Salt, T. E., and Sillito, A. M. (2006). Functional alignment of feedback effects from visual cortex to thalamus. *Nat. Neurosci.* 9, 1330–1336. doi: 10.1038/nn1768

Wiltschut, J., and Hamker, F. H. (2009). Efficient coding correlates with spatial frequency tuning in a model of V1 receptive field organization. *Visual Neurosci.* 26, 21–34. doi: 10.1017/S0952523808080966

Zhu, M., and Rozell, C. J. (2013). Visual nonclassical receptive field effects emerge from sparse coding in a dynamical system. *PLoS Comput. Biol.* 9:e1003191. doi: 10.1371/journal.pcbi.1003191

Zylberberg, J., Murphy, J. T., and DeWeese, M. R. (2011). A sparse coding model with synaptically local plasticity and spiking neurons can account for the diverse shapes of V1 simple cell receptive fields. *PLoS Comput. Biol.* 7:e1002250. doi: 10.1371/journal.pcbi.1002250

rSK1 in Rat Neurons: A Controller of Membrane rSK2?

Eleonora Autuori, Petra Sedlak, Li Xu, Margreet C. Ridder, Angelo Tedoldi and Pankaj Sah*

Queensland Brain Institute, The University of Queensland, Brisbane, QLD, Australia

In mammalian neurons, small conductance calcium-activated potassium channels (SK channels) are activated by calcium influx and contribute to the afterhyperpolarization (AHP) that follows action potentials. Three types of SK channel, SK1, SK2 and SK3 are recognized and encoded by separate genes that are widely expressed in overlapping distributions in the mammalian brain. Expression of the rat genes, rSK2 and rSK3 generates functional ion channels that traffic to the membrane as homomeric and heteromeric complexes. However, rSK1 is not trafficked to the plasma membrane, appears not to form functional channels, and the role of rSK1 in neurons is not clear. Here, we show that rSK1 co-assembles with rSK2. rSK1 is not trafficked to the membrane but is retained in a cytoplasmic compartment. When rSK2 is present, heteromeric rSK1-rSK2 channels are also retained in the cytosolic compartment, reducing the total SK channel content on the plasma membrane. Thus, rSK1 appears to act as chaperone for rSK2 channels and expression of rSK1 may control the level of functional SK current in rat neurons.

Keywords: spike frequency adaptation, potassium channel, afterhyperpolarization, excitability, calcium activated K$^+$ channels (KCa1–KCa5)

INTRODUCTION

Potassium channels are widely expressed in the central nervous system (CNS) where they play an important role in regulating the intrinsic excitability of neurons. A subset of potassium channels expressed in central neurons are Ca^{2+}-activated K^+ channels that regulate cellular excitability and spike frequency adaptation (Coetzee et al., 1999; Adelman et al., 2012). These are divided into three families that comprise the large conductance (BK) channels (KCa1.1), small conductance (SK) channels KCa2.1, KCa2.2, KCa2.3 (SK1, SK2, and SK3), and the intermediate conductance (IK) channels KCa3.1 (Vergara et al., 1998). In neurons, these channels are generally driven by calcium influx during action potentials, and activation of BK currents contributes to spike repolarization, while SK channel activity is slower, contributing to the afterhyperpolarization (AHP) that follows (Sah, 1996). These channels can also be activated by calcium release from intracellular stores again leading to inhibition of neural activity (Fiorillo and Williams, 1998), and are also expressed at glutamatergic synapses, where they are activated by calcium influx during synaptic activity and play a role in tuning synaptic plasticity (Faber et al., 2005; Ngo-Anh et al., 2005; Lin et al., 2008).

SK channels are encoded by three genes, SK1, SK2 and SK3 (Köhler et al., 1996; Stocker and Pedarzani, 2000; Sailer et al., 2002), which are expressed throughout the mammalian brain in distinct but partially overlapping distributions. In the rodent brain, SK1 and SK2 are generally co-expressed while SK3 channels are present in a complementary distribution (Stocker and Pedarzani, 2000). Functional studies in heterologous expression systems have shown that rat SK2 (rSK2) and SK3 (rSK3), and human SK1 (hSK1) channels form functional homomeric channels

that are voltage-insensitive and gated by the binding of Ca^{2+} to calmodulin (CaM; Xia et al., 1998; Lee and MacKinnon, 2018), which is covalently linked at their cytosolic, carboxy-terminal region (Lee and MacKinnon, 2018). In contrast, the rat SK1 (rSK1), does not generate functional homomeric channels (Köhler et al., 1996; D'Hoedt et al., 2004), but does appear to co-assemble with rSK2 (Ishii et al., 1997; Benton et al., 2003; Church et al., 2015). This difference is due to differences in the sequence identity of the two channels. Thus, rSK2 and rSK3 are highly homologous to the human SK2 (97.6%) and SK3 (94.4%), and have similar pharmacology and functional profiles (Köhler et al., 1996; Joiner et al., 1997; Desai et al., 2000), while rSK1 is only 84% homologous to hSK1 (D'Hoedt et al., 2004). Indeed, replacing the carboxy terminal of rSK1 with that of rSK2 restores functional expression of rSK1, and swapping C- and N-termini of hSK1 with those from rSK1 prevents expression of functional hSK1 channels on the cell membrane of HEK293T cells (D'Hoedt et al., 2004).

In summary, while rodent rSK1, rSK2 and rSK3 channels are widely expressed in neurons of the rodent nervous system, rSK1, unlike rSK2 and rSK3, does not form functional homomeric channels. Homomers and dimers consisting of SK2 and SK3 form functional channels and contribute to the AHP and are also involved in synaptic plasticity (Adelman et al., 2012). It has been reported (Benton et al., 2003) that co-expression of rSK1 and rSK2 in HEK293 cells results in larger calcium-activated currents with altered pharmacology suggesting that rSK1 can form functional heteromeric assemblies with rSK2, though biochemical evidence for this co-assembly is lacking. Thus, while rSK1 is widely co-expressed with rSK2, the functional role of rodent SK1 remains unclear. In this study, we test the function of rSK1 channels in a heterologous system and cultured rat hippocampal pyramidal neurons. Firstly, we tested if rSK1 and rSK2 could interact in a heterologous system. We then looked at the effect of either overexpressing or downregulating rSK1 channels and tested the level of rSK2 channels expressed on the cells membranes of both a heterologous system and in hippocampal pyramidal neurons. Finally, we tested if changing rSK1 levels was related to a change in medium AHP (mAHP) and its underlying current (I_{AHP}) in hippocampal pyramidal neurons that are generated by the activation of SK2 channels (Stocker et al., 1999, 2004).

MATERIALS AND METHODS

Neuronal Cultures
Primary E18 Wistar rat hippocampal cultures were prepared as previously described (Delaney et al., 2013). Briefly, hippocampi were digested for 20 min at 37°C in papain (12 U/ml, Worthington, suspension 28.4 U/mg) made up in dissection medium (1× HBSS, 1% penicillin/streptomycin, 1% pyruvate, 10 mM HEPES, 30 mM glucose) and supplemented with 1% DNase (Sigma). The digested material was washed three times in plating medium [Neurobasal medium (Life Technologies) supplemented with 5% heat-inactivated fetal bovine serum (FBS), 1% penicillin/streptomycin, 1% Glutamax (Life Technologies) and 2% B27 (Life Technologies)] and the resulting pellet was

triturated in plating medium using a fire-polished glass Pasteur pipette. Cells were plated at a density of 1×10^5 cells/ml for immunocytochemistry and the biotinylation assays and at 4×10^4 cells/ml for the electrophysiology experiments, on glass coverslips precoated with poly-D-lysine in plating medium and maintained at 37°C in 5% CO_2.

Biotinylation and Western Blot Analysis
Hippocampal neurons were cooled on ice and were then rinsed three times with phosphate buffered saline (PBS) prior to exposure with EZ-Link Sulfo-NHS-SS-Biotin (Pierce), dissolved in PBS to a concentration of 1.22 mg/ml. The biotinylation reaction was undertaken on ice for 30 min, followed by biotin removal and washing twice with 100 mM glycine in PBS and then once with PBS. Cells were lysed in 200 µl lysis buffer (20 mM sodium phosphate, pH 7.5, 150 mM NaCl, 0.5% NP40, 0.5% sodium deoxycholate, 0.1% SDS, 1× complete protease inhibitors, Roche) and left on ice for 30–45 min. The lysates were then centrifuged at 6,500 g for 5 min. Twenty microliters of the resulting supernatant was then removed and this was classified as the total lysate (LYS). Biotinylated proteins were captured from the supernatant by the addition of 25 µl Streptavidin Agarose Resin (Pierce) at 4°C for 30 min. The beads were then centrifuged at 10,000 g for 1 min and the resulting supernatant was classified as the cytoplasmic phase (CYTO). The beads were washed a further three times in PBS. This was classed as the membrane fraction (MEMB). All samples were then boiled for 5–10 min in 1× SDS sample buffer containing 100 mM DTT and then fractionated on 4%–12% Bis-Tris gradient gels (Life Technologies) and transferred to polyvinylidene difluoride membrane (Immobilon-P, Merck Millipore) at 150 V in 1× MOPS transfer buffer (Life Technologies). Blots were blocked in Tris-buffered saline (TBS) containing 5% skim milk powder, probed with primary antibody [mouse α Myc (9B11; 1/5,000, Cell Signalling Technology), rabbit α HA (1/500, Cell Signalling Technology), mouse α β-actin (AC-15; 1/20,000, Cell Signalling Technology), rabbit α EGFR (1/2,000, Cell Signalling Technology), mouse α Na$^+$/K$^+$ ATPase (alpha 1; clone C464.6; 1/4,000, Merck Millipore), rabbit α SK2 (c-39; 1/2,000, gift of J. Adelman)] followed by incubation with horseradish peroxidase-conjugated goat anti-rabbit or mouse IgG (1/20,000, Biorad) and detection by SuperSignal West Pico or FEMTO chemiluminescent substrate (Pierce). Blots were scanned using the Odyssey Infrared Imaging System (Li-cor) and densitometry analysis was carried out using Image Studio Lite software.

All data are expressed as mean ± standard error of the mean (SEM). Statistical analysis was performed using GraphPad Prism (GraphPad Software). All the data were tested for normal distribution using a normality test. If data were normally distributed, a Student's unpaired t-test was used. If the data did not pass the normality test, the Mann-Whitney test was used. Significance was determined at $p < 0.05$.

Co-immunoprecipitation
Supernatants were precleared for 1 h at 4°C with 75 µl of 50% Sepharose bead slurry (Amersham) and 1.5 µg of species-

specific IgG (Sigma, 1 mg/ml). Protein G sepharose bead slurry and mouse IgG were used for the samples immunoprecipitated with mouse α Myc (9B11; Cell Signalling Technology); protein A sepharose bead slurry and rabbit IgG for the samples immunoprecipitated with the rabbit α HA (Cell Signalling Technology) antibody. Samples were then spun at 13,000 rpm for 5 min and the resulting supernatants were incubated overnight at 4°C respectively with mouse α Myc antibody (2 μl/1.5 mg proteins) or rabbit α HA (1:50) antibodies. The following day, the α Myc samples were incubated with 75 μl of 50% protein G sepharose beads and the α HA samples with 75 μl of 50% protein A sepharose beads for a further 3 h at 4°C. Samples were centrifuged at 3,000 rpm for 5 min at 4°C and the recovered pellets were washed three times in RIPA buffer [150 mM NaCl, 1% Triton X-100, 0.5% sodium deoxycholate, 0.1% SDS, 50 mM Tris pH = 8.0 and protease inhibitors (Roche)]. Samples were eluted with sample buffer containing DTT and denatured by boiling for 5 min. Co-immunoprecipitation samples (IP) were run together with the input (IN) and the supernatant (S) resulting from the immunoprecipitation, transferred as described above and blotted with mouse α Myc and rabbit α HA as described above.
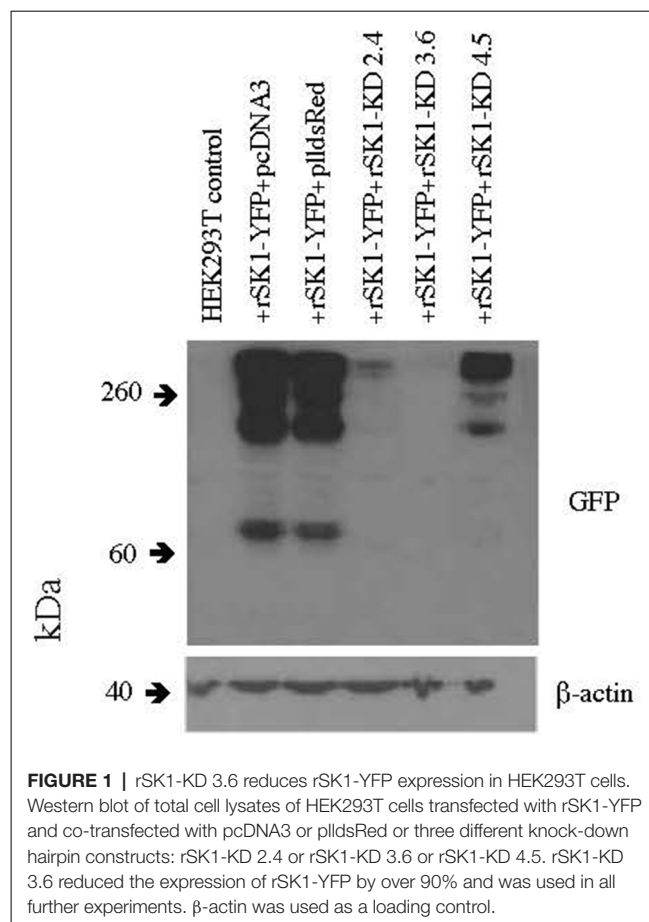
Lentiviral Constructs

Several knock-down hairpin constructs were produced using annealed phosphorylated oligonucleotides, which were ligated into the pll3.7 dsRed vector digested with *HpaI* and *XhoI*. The pll3.7 dsRed vector has the EGFP of pll3.7 vector replaced with dsRed. HEK293T cells were plated at a density of 5×10^5 cells per 35 mm well and were transfected 2 h later using Lipofectamine 2,000 (Life Technologies) as per the manufacturer's instructions with constructs expressing rat SK1-YFP (gift of G. Moss) and three different plasmids carrying SK1 knock down hairpins (rSK1-KD 2.4, 3.6 and 4.5) at a ratio of 1:1. Forty-eight hours later, the cells were checked for YFP and dsRed expression, washed with PBS and lysed in 600 μl 1× sample buffer. Thirty microliters of this was run on an SDS-PAGE gel, transferred and western blots carried out using rabbit α GFP (1/500, Merck Millipore) primary antibody as per methods above except that x-ray film was used for detection. Mouse β-actin (AC-15; 1/20,000, Cell Signalling Technology) was used as a loading control. Out of the three plasmids, rSK1-KD 3.6 was the most successful at reducing SK1-YFP expression (**Figure 1**). The primers used to produce this knockdown construct are as follows–

- pLLrSK1_3S: tgtctcatagcccaagccatatttcaagagaatatggcttgggctatgagactttttc
- pLLrSK1_3AS: tcgagaaaaaagtctcatagcccaagccatattctcttgaaatatggcttgggctatgagaca

To enable us to determine the virally infected neurons, we replaced the CMV promoter (expressing the dsRed) in pll3.7 dsRed with the hSyn promoter using the Not and Nhe restriction sites and the following primers–

- Synapsin400-Not-F: aaagcggccgcgtgtctagactgcagagggccct
- Synapsin400-Nhe-R: aaagctagcttctcgactgcgctctcagg

FIGURE 1 | rSK1-KD 3.6 reduces rSK1-YFP expression in HEK293T cells. Western blot of total cell lysates of HEK293T cells transfected with rSK1-YFP and co-transfected with pcDNA3 or plldsRed or three different knock-down hairpin constructs: rSK1-KD 2.4 or rSK1-KD 3.6 or rSK1-KD 4.5. rSK1-KD 3.6 reduced the expression of rSK1-YFP by over 90% and was used in all further experiments. β-actin was used as a loading control.

A scrambled (SCRAM) control plasmid was also produced using a random combination of the hairpin sequence of the successful knockdown construct. The primers used to produce the SCRAM construct are as follows–

- SCRAM pLLrSK1_3S: tggcactttaacgccaatcactttcaagagaagtgattggcgttaaagtgcctttttc
- SCRAM pLLrSK1_3AS: tcgagaaaaaaggcactttaacgccaatcacttctcttgaaagtgattggcgttaaagtgcca

The rat SK2 overexpression plasmid pJPA5.rSK2.Myc3 (gift of J. Adelman) has a triple Myc tag located extracellularly between the S3 and S4 domains. The Myc-tagged SK2 was further cloned into the lentiviral FUGW plasmid (gift of P. Osten) using the AgeI and EcoRI restriction sites and the following primers–

- F-rSK2: gggaccggtgaaatagccatgagcagctgcaggtacaacggg
- R-rSK2: ggggaattcgctactctcagatgaagttgg

The rat SK1 overexpression plasmid pcDNA3.1-rSK1_HA (gift of G. Moss) has a HA tag located intracellularly between the S2 and S3 domains. The HA-tagged SK1 was further cloned into the lentiviral FUGW plasmid (gift of P. Osten).

For lentivirus production, all plasmids (pMDG, pMDL g/p RRE, pRSV Rev and the pll3.7 transfer vector) were prepared using the Qiagen Endofree maxiprep kit and lentivirus was prepared via calcium phosphate transfection of 80%

confluent triple T175 flasks of HEK293T cells. Cells were grown in DMEM (Life Technologies) plus 10% FBS and 1% penicillin/streptomycin (Life Technologies). Forty microgram of transfer vector together with 20 µg of the other three plasmids were transfected per triple T175 flask. Plasmids were diluted to 0.5 µg/ml in TE buffer and made up to a volume of 3 ml with sterile distilled water. Three-hundred microliter of 2.5 M CaCl$_2$ was added together with 3 ml of 2× BBS (50 mM BES, 280 mM NaCl, 1.5 mM Na$_2$HPO$_4$, pH = 6.95). This was incubated for 20 min at room temperature before adding the mix dropwise to the cells. The cells were incubated for 3–4 h at 37°C plus 5% CO$_2$ before the removal of the calcium phosphate mix and addition of new media. Viral supernatants were removed 48 h post-transfection, filtered using a 0.45 µm filter unit and then centrifuged for 4 h at 20,000 g over a 20% sucrose cushion. The resulting pellet was resuspended overnight in PBS containing 1% BSA and then ultracentrifuged at 44,000 g for 90 min to concentrate the virus. The virus pellet was resuspended in PBS containing 1% BSA.

Neuronal cultures were infected at 4–5 DIV and the virus was left on for 4–12 h. The virus was then removed and conditioned media added to the cultures. Biotinylation experiments and quantitative PCR (qPCR) were undertaken 4–5 days post infection. Knockdown and overexpression lentiviral constructs were also tested for specificity and expression using qPCR. Briefly, RNAs were purified using a RNeasy kit (Qiagen), DNase 1 (Qiagen) treated and reverse transcribed to produce complementary DNA using random hexamers and the SuperScript III First Strand Synthesis System (Life Technologies), as per the manufacturer's instructions. Standard qPCR was carried out on a Rotorgene RG-3000 Thermocycler using the Platinum SYBR Green qPCR UDG SuperMix (Life Technologies) and 0.2 µM of the following primers:

- rSK1 Fwd—tca tct cca tta cct tcc tg
- rSK1 Rev—agc ctg gtg tgt ttg tag at
- rSK2 Fwd—gtc gct gta ttc ttt agc tct g
- rSK2 Rev—acg ctc ata agt cat ggc
- rGAPDH Fwd—gag tct act ggc gtc ttc ac
- rGAPDH Rev—cca tcc aca gtc ttc tga gt

Reactions were performed in either duplicate or triplicate. The PCR program was as follows: 50°C 2 min, 95°C 10 min, 35 cycles −95°C 10 s, 54°C 15 s, 72°C 20 s. Relative expression of rSK1 and rSK2 mRNA was obtained using the comparative Ct method.

Cosm6 Transfections

For the Cosm6 transfections- pJPA5.rSK2.Myc3 and pcDNA3.1-rSK1_HA along with pcDNA3 for the single construct transfections were transfected using the same protocol as described above but using a ratio of 1:2:1.

Immunohistochemistry

Hippocampal cultures were washed in PBS and fixed in 4% PFA containing 4% sucrose. After three washes in PBS, the cells were first blocked in a 0.3% BSA/0.1% Triton X-100 solution, then incubated in primary antibody overnight at 4°C, washed in PBS

and left in secondary antibody (Life Technologies) for a further 1 h at room temperature. Stained cultures were analyzed on a confocal fluorescence microscope (Zeiss LSM510).

Electrophysiology

Rat hippocampal neuronal cultures were recorded at least 2 weeks after viral infection (DIV 15–17). Coverslips were transferred from the incubator to the recording chamber, located on a stage of a fluorescence microscope (Zeiss 710). Ringer solution (140 mM NaCl, 5 mM KCl, 2 mM CaCl$_2$, 1 mM MgCl$_2$ 10 mM glucose and 10 mM HEPES) was constantly perfused at 1–1.5 ml/min at room temperature. Whole-cell recordings were obtained from dsRed-positive neurons (SCRAM or SK1 KD) or neurons (rSK1-HA and rSK2-myc) using a K-methyl-sulfate based internal solution containing (mM): 135 KMeSO$_4$, 5 NaCl, 10 Hepes, 2 Mg$_2$-ATP, 0.3 Na$_3$-GTP, 0.1 spermine and 7 phosphocreatine (pH = 7.3, ∼290 mOsmol).

Tetrodotoxin (TTX, 1 µM, Sigma), D-(−)-2-Amino-5-phosphonopentanoic acid (D-APV, 20 µM, Tocris) and 6-cyano-7-nitroquinoxaline-2,3-dione disodium salt (CNQX, 20 µM, Tocris) were added to the ringer solution to reduce spontaneous activity and isolate mAHP and I_{AHP}. For a recording of mAHP, cells were injected with a 400 pA/100 ms step either at resting membrane potential or by clamping the cells at −50 mV. For recordings of I_{AHP}, cells were held at −50 mV and a depolarizing step (+60 mV) to +10 mV for 100 ms was used to increase Ca^{2+} influx through voltage-gated Ca^{2+} channels, which are necessary to activate I_{AHP} (Stocker et al., 2004).

Data were collected using Axograph X software and a Multiclamp 700B amplifier (Molecular Devices). Signals were filtered at 10 kHz and digitized at 50 kHz using an ITC-16 A/D converter (InstruTech).

RESULTS

rSK1 and rSK2 Co-assemble in Cosm6 Cells

To test expression of SK channels, rSK1 and rSK2 were epitope-tagged with HA and Myc, respectively (see "Materials and Methods" section). As shown previously (D'Hoedt et al., 2004; Church et al., 2015), transfection in Cosm6 cells shows that rSK2 was expressed throughout the cell, while rSK1-HA was restricted to the cytosolic somatic compartment (**Figure 2**), likely the endoplasmic reticulum and Golgi (Church et al., 2015). As shown by others (Strassmaier et al., 2005), western blot analysis using Myc and HA antibodies shows bands of several sizes that correspond to monomeric protein (∼60 kDa, band 3), dimers (∼120 kDa, band 2) and a large molecular band (∼200 kDa, band 1), which is likely to be a tetramer (**Figure 2B**). Separation of the membrane fraction (MEMB on blots) using surface biotinylation (see "Materials and Methods" section) shows that unlike rSK2, rSK1 channels are not detectable on the plasma membrane. When rSK1 and rSK2 were co-expressed, again rSK2 was detectable in the membrane fraction but rSK1 was not (**Figure 2B**), but the level of rSK2 protein in the membrane fraction was reduced (**Figures 2B,C**). There was a clear reduction
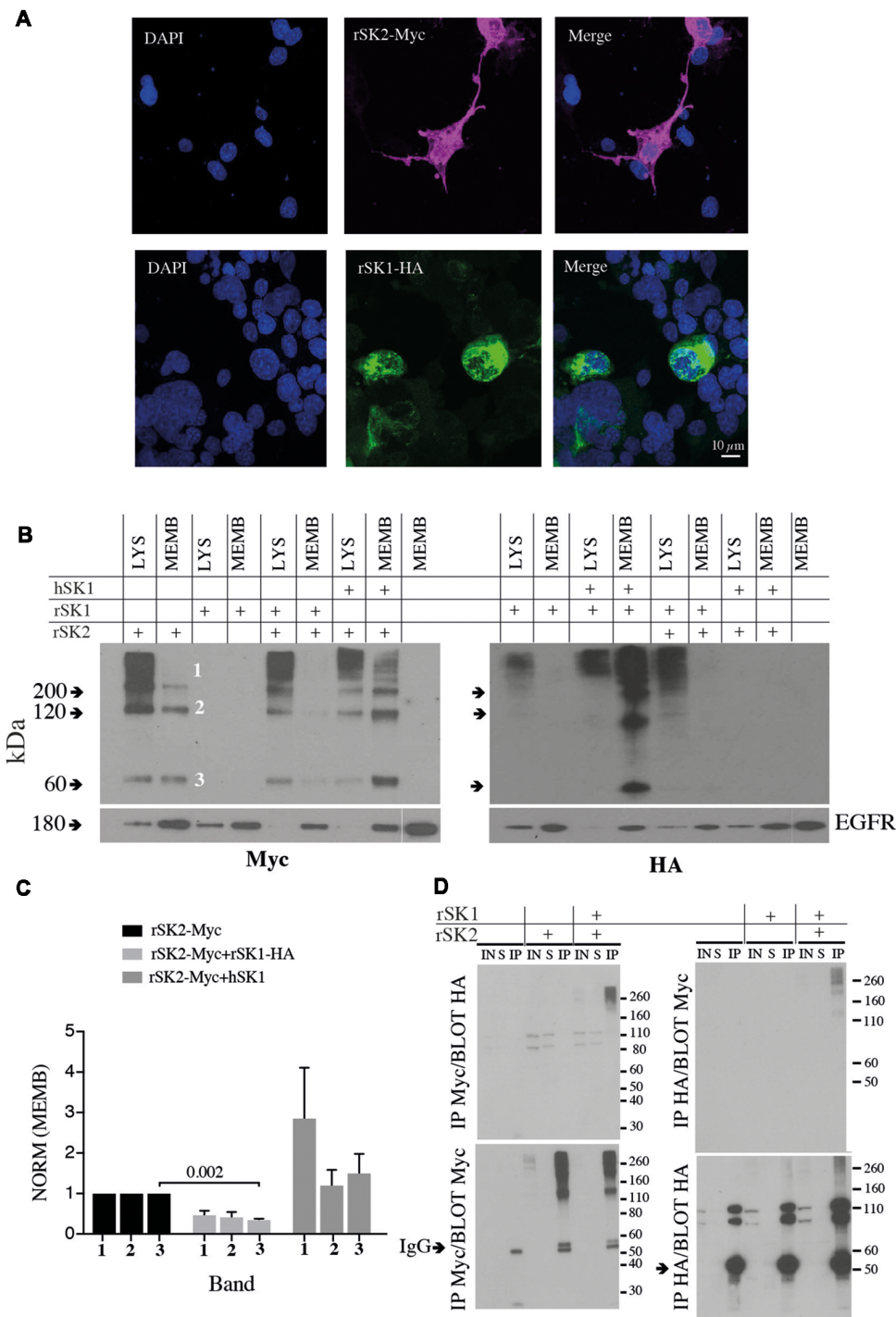
FIGURE 2 | rSK1 regulates the membrane expression of rSK2 channels. **(A)** Immunocytochemistry of Cosm6 cells transfected with rSK2-Myc or rSK1-HA. rSK2-Myc (pink), rSK1-HA (green), DAPI (blue). **(B)** Immunoblots from Cosm6 cells transfected with rSK1-HA, rSK2-Myc and hSK1 in different combinations. Shown are the total lysate (LYS) and membrane (MEMB) fraction isolated using surface biotinylation probed with either Myc (rSK2; left) or HA (rSK1; right) antibody. EGFR was used as the loading control. rSK2 transfection results in three size bands that represent monomeric protein (band 3, ~60 kDa), dimers (band 2, ~120 kDa) and tetramers (band 1, >200 kDa). Co-transfection of rSK2 and rSK1 reduced the amount of rSK2 expressed in the cell membrane. **(C)** Bar graphs show normalized relative density of SK2, as described in the "Materials and Methods" section, in the membrane fraction from Cosm6 cells transfected with rSK2 alone or co-transfected with rSK1 or hSK1. Co-transfection with rSK1-HA significantly reduced the 60 kDa band ($p = 0.002$). Co-transfection with hSK1 increased expression of rSK2. **(D)** Co-immunoprecipitation (co-IP) assay of transfected Cosm6 cells shows co-assembly of rSK1-rSK2. Solubilized protein from Cosm6 cells transfected with rSK1 and/or rSK2 was immunoprecipitated with Myc (left) or HA (right) and immunoblotted with either Myc or HA as indicated.

in monomeric rSK2 ($n = 3$, $p = 0.002$), but other higher molecular weight species were also greatly reduced (**Figure 2C**). In contrast to rSK1, hSK1 is known to express as functional channels (D'Hoedt et al., 2004), and forms heteromultimeric channels (Ishii et al., 1997; Benton et al., 2003; Church et al., 2015). In agreement, co-expression of hSK1 with the rSK2 increased the amount of rSK2-Myc on the cell membranes (**Figures 2B,C**).

What explains the reduction in membrane rSK2 when it is co-expressed with rSK1? As rSK1 is retained in cytosolic compartments and not trafficked to the membrane (Benton et al., 2003; Church et al., 2015), it is possible that rSK2 co-assembles with rSK1 with the resulting heteromultimers being retained, leading to a reduction of membrane rSK2. We, therefore, tested for interactions between these two channels using co-immunoprecipitation (co-IP). rSK1-HA and rSK2-Myc were expressed in Cosm6 cells, and Myc or HA antibodies used to isolate rSK2 or rSK1 using sepharose A/G beads followed by immunoblot analysis of SK protein. IP of Cosm6 cell homogenates using anti-Myc antibody revealed the presence of rSK2 in both supernatant, input and co-immunoprecipitated samples, from single transfected (rSK2-Myc) and co-transfected cells (rSK2-Myc+rSK1-HA; **Figure 2D**). Probing with HA antibody revealed the presence of rSK1-HA in the multimeric complex (**Figure 2D**, left). Complimentary experiments using the anti-HA antibody to precipitate rSK1 protein lysates and probing with anti-Myc also revealed the presence of rSK2 in the multimeric high molecular weight band (**Figure 2D**, right). It was notable that following IP with myc and probe with HA or the reverse, while very high molecular weight (>200 kD) complexes are present, there is a dearth of complexes at ~110 kD (**Figure 2D**, upper panels). We interpret this to show the presence of tetrameric heteromeric protein.

Overexpression of rSK1 in Neurons Decreases Plasma Membrane rSK2

These results show that when expressed in Cosm6 cells rSK1 and rSK2 can co-assemble. Moreover, the interaction between the two channels affects the total rSK2 present in the cell membrane. Thus, rSK1 channels appear to "trap" rSK2 in the cytoplasmic compartment, therefore reducing the total amount of rSK2 trafficked to the cell membrane.

To test if rSK1 acts similarly in neurons, we turned to cultured rat hippocampal neurons. rSK1-HA and rSK2-Myc were delivered using lentivirus (see "Materials and Methods" section). We first tested the effect of overexpression of rSK1 or rSK2 on mRNA levels, using qPCR. Increasing the volume of rSK1 or rSK2 expressing virus added to neuronal cultures increased the relative expression of mRNA of the specific gene delivered without affecting the other: infection of 1×10^5 rat hippocampal cultures with 0.2 µl and 0.5 µl rSK1-HA lentivirus increased rSK1 relative expression by 60 and 200-fold, respectively, while infection with 0.2 µl and 0.5 µl of rSK2-Myc lentivirus increased rSK2 relative expression by 10 and 40-fold, respectively (data not shown). Next, expression of SK2 protein was tested using immunocytochemistry (**Figure 3**). Transduction of rSK2-Myc (**Figure 3A**) or rSK1-HA (**Figure 3B**) showed clear expression of Myc and HA in cells positive for the neuronal marker NeuN.

Co-expression of rSK1 and rSK2 show that the two are at least partially co-localized in neurons (**Figure 3C**), consistent with heterodimer formation of rSK1 and rSK2. It can also be seen that in neurons when rSK1-HA is expressed alone (**Figure 3B**), its distribution is not confined to the soma, but spreads throughout the cell, consistent with the expression of endogenous SK2 in these neurons (**Figure 3D**).

To test for the presence of rSK2 in the plasma membrane we again turned to the surface biotinylation assay to separate the membrane compartment. rSK1-HA and rSK2-Myc were transduced using lentivirus. As in Cosm6 cells (**Figure 2**), we found that co-expression of rSK2 with rSK1 decreased the amount of rSK2 detected in the membrane fraction of infected neurons (**Figures 4A,B**). This reduction was particularly significant in the high molecular weight band, again suggesting delivery of a heteromultimeric protein ($n = 3$, $p = 0.04$; **Figure 4B**). To account for possible effects of overexpressing rSK2, we tested the impact of rSK1 expression on endogenous rSK2 using the specific SK2 antibody (c-39). When rSK1-HA was transduced in hippocampal neurons, the endogenous levels of rSK2 protein detected in the cell membrane was reduced (**Figure 4C**, upper panel), while the cytoplasmic fraction increased (**Figure 4C**, bottom panel). Quantification of the relative density of endogenous rSK2 (c-39) normalized to β-actin shows that the density of the high molecular band was significantly reduced in the membrane fraction in comparison to uninfected neurons ($n = 5$, $p = 0.001$; **Figure 4D**), but was increased in the cytosolic fraction ($n = 4$, $p = 0.006$; **Figure 4D**, bottom panel). This increase in SK2 in the "cytoplasmic" fraction is likely due to the fact that rSK1-rSK2 heteromultimers are not trafficked to the plasma membrane, but remain trapped in the cytoplasmic compartment, possibly the endoplasmic reticulum.

rSK1 Modulates Functional SK Channels

In mammalian neurons, SK channels contribute to the medium duration AHP that follows action potentials (I_{AHP}), and in some types of neurons they are also present at excitatory synapses where they modulate synaptic strength (Faber et al., 2005; Ngo-Anh et al., 2005; Adelman et al., 2012). We have shown that overexpression of rSK1 seems to reduce the amount of membrane rSK2 protein. We next tested if altering endogenous rSK1 levels would have an impact on functional membrane rSK channels in neurons. To disrupt endogenous rSK1, we used short hairpin RNA interference (RNAi) constructs to reduce rSK1 expression in hippocampal neurons using lentivirus (see "Materials and Methods" section). rSK2 protein levels in the membrane were assessed using the biotinylation assay and western blot analysis (**Figures 5A,B**). Knockdown of rSK1 (rSK1-KD) significantly reduced the amount of rSK2 monomers (60 kDa) expressed in the cell membrane in comparison to neurons infected with scrambled virus ($n = 4$, $p = 0.008$), but the total membrane rSK2 tetramers (>200 kDa) appeared to be higher than the scrambled control (**Figure 5A2**). In the cytosolic compartment, high molecular weight (>200 kDa) rSK2 protein was significantly reduced as compared to scrambled control cells ($n = 4$, $p = 0.03$; **Figure 5B2**).
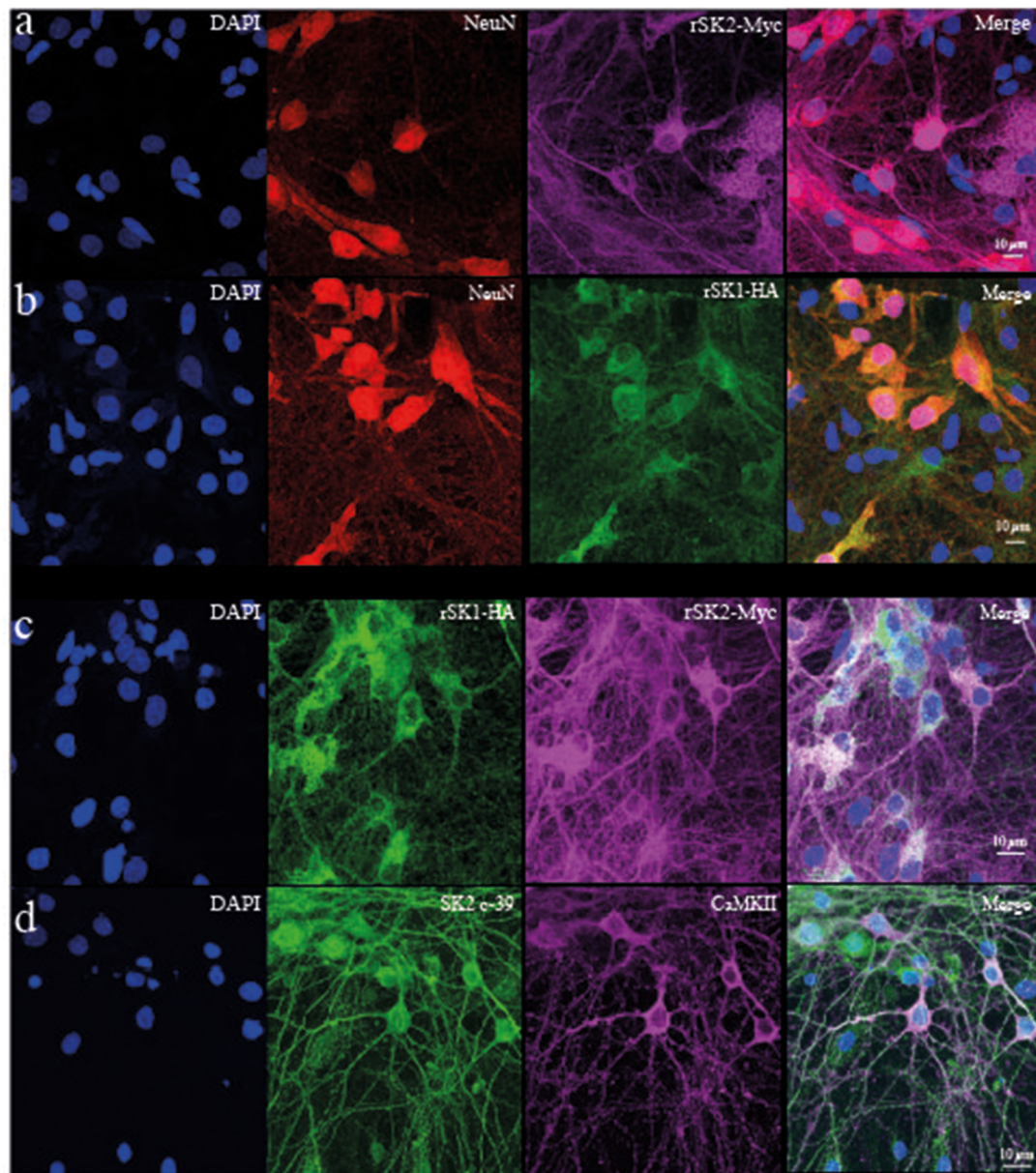
FIGURE 3 | rSK1 and rSK2 expressed in rat hippocampal neuronal cultures. Cultured hippocampal neurons were transduced with rSK2-Myc **(A)** or rSK1-HA **(B)** or both **(C)** using lentivirus and immunostained for DAPI (blue), the neuronal marker NeuN (red), Myc (purple) or HA (green). **(A,B)** SK2 (Myc) and SK1 (HA) are expressed throughout the soma and processes of transduced neurons. DAPI **(C)** co-expression of rSK2 and rSK1 shows that both proteins are present in the same compartments. **(D)** Immunocytochemistry using a specific SK2 antibody (c-39; green) shows that endogenous rSK2 is present in the soma and throughout the processes of cultured neurons. Hippocampal neurons are marked using CaMKII (purple).

To test for functional expression of rSK channels, whole-cell recordings were obtained from infected hippocampal neurons in culture (**Figures 5C,D**). The passive membrane properties of neurons in infected with SK2 RNAi hairpins and scrambled controls were not different, and are given in **Supplementary Table S1**. Reducing rSK1 with RNAi increased the amplitude of I_{AHP} in comparison to that in scrambled infected neurons ($p = 0.03$; **Figure 5C**). As expected, switching to current clamp also revealed an increase in the amplitude of the medium AHP

($p = 0.03$; **Figure 5D**), and this enhanced I_{AHP} current was blocked by apamin (**Figure 5E**). Together, these results support a role for rSK1 channels in regulating the expression of rSK2 on the cell membrane.

DISCUSSION

SK1, SK2 and SK3 encode calcium-activated potassium channels that are widely expressed in the mammalian brain. In
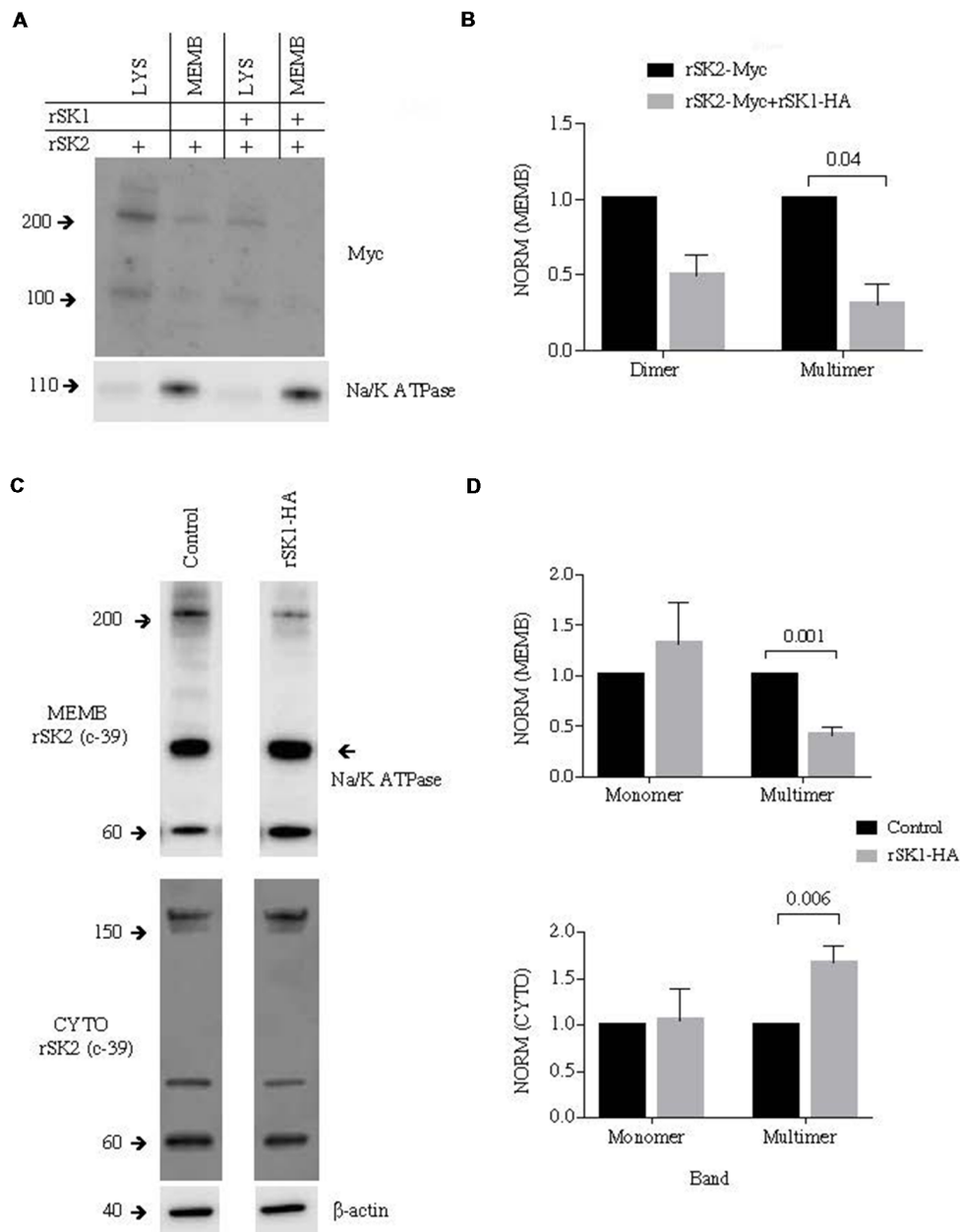
FIGURE 4 | rSK1 regulates membrane expression of rSK2 channels in cultured neurons. Cultured hippocampal neurons were transduced to express rSK1-HA or rSK2-Myc or both. Surface protein was biotinylated, and solubilized protein from neurons separated into total lysate (LYS) and membrane fraction (MEMB) as detailed in the "Materials and Methods" section. Immunoblots were probed using the Myc antibody to test for rSK2 expression. Na/K ATPase or β-actin was used as loading controls. **(A)** Co-expression of rSK1 and rSK2 reduces membrane expression of rSK2. **(B)** Graph shows normalized relative density of SK2, in the membrane fraction from neurons transduced with rSK2 alone or rSK2 and rSK1. Co-infection significantly reduced the band at 200 kDa (p = 0.04). **(C)** Expression of rSK1 in neurons reduced surface expression of endogenous rSK2. Cultured neurons were transduced to express rSK1 using lentivirus. Blots show solubilized protein from the membrane fraction (upper blots) and cytosolic fraction (lower blots) from control neurons (left) and neurons transduced with rSK1(right). Blots were probed with anti SK2 antibody (c-39). **(D)** Quantification of SK2 protein expression in the membrane (upper graph) and cytosolic (lower graph) fraction from control neurons and neurons transduced with rSK1. Neurons infected with rSK1 had a significantly lower amount of endogenous high molecular weight rSK2 in the membrane fraction and this was balanced by an increase in the total cytoplasmic fraction.

rodents, rSK1 and rSK2 are expressed in strongly overlapping distributions. However, while expression of rSK2 channels produces functional calcium-activated potassium channels, rSK1 channels are made but not trafficked to the plasma membrane (D'Hoedt et al., 2004; Church et al., 2015), and the functional role of these channels is not clear. In this study, we show that rSK1 channels co-assemble with rSK2, and regulate the plasma membrane levels of rSK2. Thus, expression of rSK1 in
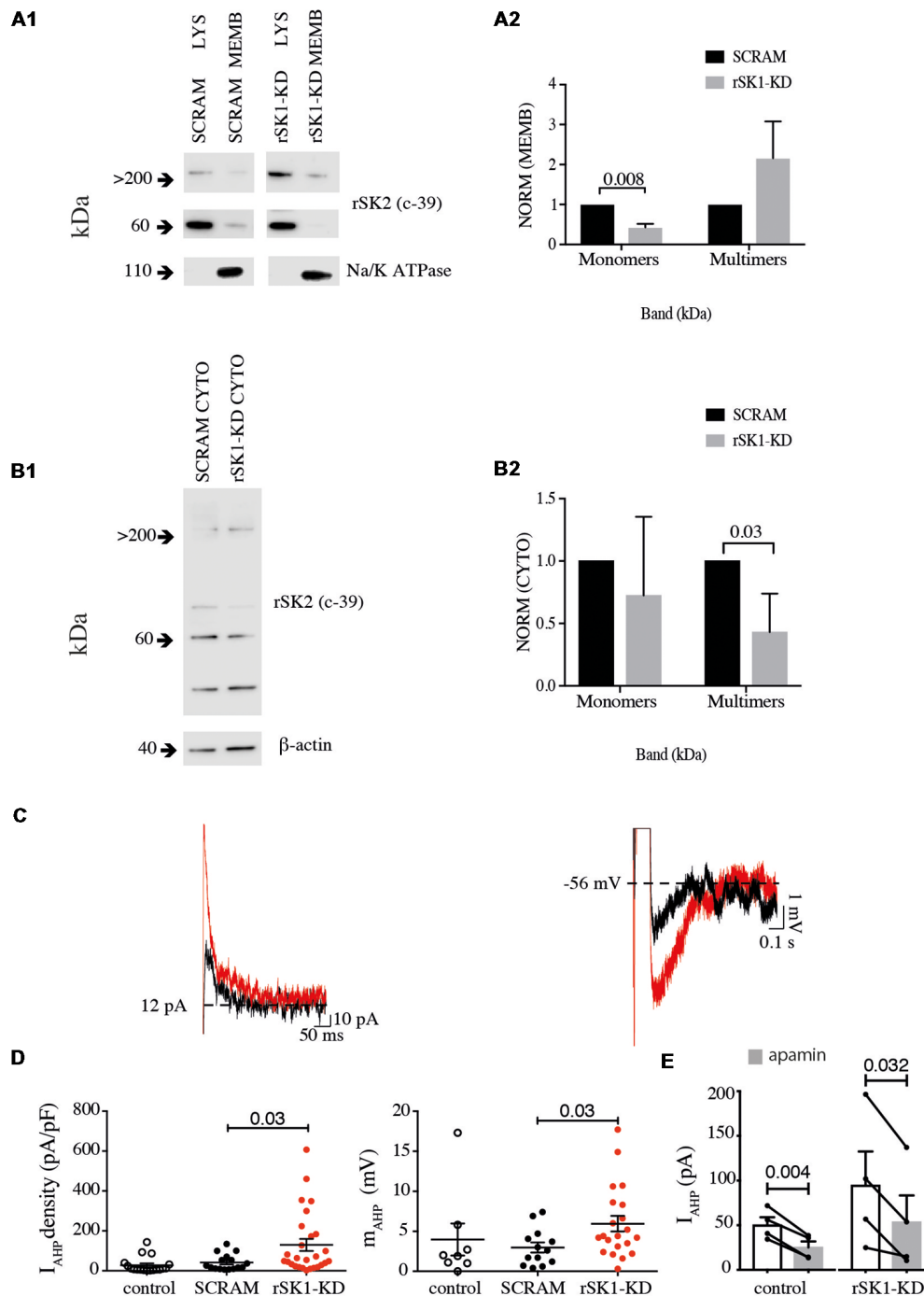
FIGURE 5 | Knock down of rSK1 enhances I$_{AHP}$ in hippocampal neurons. Cultured hippocampal neurons were transduced to express a hairpin knockdown vector for SK1 (rSK1-KD) or a scrambled control vector (SCRAM). Surface protein was biotinylated, and solubilized protein from neurons separated into total lysate (LYS) and membrane fraction (MEMB) as detailed in the "Materials and Methods" section. Blots were immunoprobed using the rSK2 antibody (c-39). **(A1)** Knockdown of SK1 protein results in an overall reduction of total endogenous monomeric SK2 protein, but enhancement of high molecular weight SK channels in the plasma membrane. Na$^+$/K$^+$ ATPase was sued as the loading control. **(A2)** Normalized relative density of endogenous rSK2 in the membrane fraction [NORM (MEMB)] from neurons infected with rSK1-KD. rSK1-KD significantly decreased the amount of endogenous monomeric rSK2, but increased high molecular weight rSK2. Na$^+$/K$^+$ ATPase (~110 kDa) was used to normalize all the loaded samples for the MEMB fraction. **(B1)** Knockdown of rSK1 results in a reduction in high molecular weight SK2 in the cytoplasmic fraction. **(B2)** Quantification of blot shown in B1. β-actin (~40 kDa) was used to normalize all the loaded samples for the CYTO fraction. **(C)** Knockdown of rSK1 in hippocampal neurons (red traces) increased the amplitude of the I$_{AHP}$ in voltage clamp (left) and the medium AHP in current clamp (right) in comparison to SCRAM infected neurons (traces in gray). **(D)** Quantification of the impact of the increase in the I$_{AHP}$ current shown as pA/pF and the medium AHP following knockdown of rSK1. **(E)** The enhanced I$_{AHP}$ current following knockdown of rSK1 is apamin (100 nM) sensitive consistent with SK2 channels.

neurons leads to the formation of heteromultimers containing rSK2 that are trapped in the cytoplasmic compartment and a reduction in the total rSK2 in the plasma membrane. In support of this, knockdown of endogenous rSK1 in cultured neurons results in an increase in membrane rSK2. These results suggest that rSK1 channels, rather than functioning as ion channels, are involved in regulating the expression of rSK2 channels in the plasma membrane.

While SK channels were cloned more than 20 years ago, antibodies for immunoprecipitation and immunohistochemistry for all three members are not easily available, thus we have used epitope-tagged channels to study their expression and trafficking. Here using rSK1 tagged with HA (rSK1-HA) and rSK2 tagged with Myc (rSK2-Myc), in agreement with previous studies (D'Hoedt et al., 2004; Church et al., 2015), we show that when expressed in Cosm6 cells, rSK2 channels are present in the plasma membrane, rSK1 channels are made but not trafficked from the cytoplasmic compartment, with little or no protein detectable in the membrane compartment. When rSK1 and rSK2 are co-expressed in Cosm6 cells, we show using co-immunoprecipitation that these channels co-assemble and there is a significant reduction in the amount of rSK2 in the plasma membrane. It is well known that hSK1 channels can form functional channels that are expressed on the cell surface (D'Hoedt et al., 2004; Church et al., 2015). In agreement with this, co-expression of rSK2 with hSK1 had an opposite effect to that of rSK1 with an increase in membrane rSK2.

These results in Cosm6 cells were confirmed in rat hippocampal neurons. Co-expression of rSK1-HA and rSK2-Myc resulted in an overall reduction in the amount of rSK2-Myc trafficked to the membrane. Importantly, expression of exogenous rSK1-HA in cultured neurons also resulted in a reduction in the total endogenous rSK2, showing that expression of rSK1 can modify the levels of membrane rSK2 channel. Finally, we show that by reducing the amount of rSK1 channels using RNAi, increases the amount of endogenous rSK2 channels expressed in the cell membrane, and is consistent with the finding that transgenic mice lacking SK1 show no overall effects on hippocampal neuronal electrophysiology (Bond et al., 2004). We show that reducing the rSK1 content of hippocampal neurons reduces the overall amount of rSK2 monomeric protein in the membrane but increases the total large molecular weight fraction (**Figure 5**). This is perhaps expected as we have not modified the expression level of rSK2 protein, and as rSK1 levels are lower, more channels are trafficked to the membrane where they co-assemble as high molecular weight homomeric multimers. As a result, total monomeric protein levels are lower. This increase in membrane rSK2 levels has a functional impact as it also increases the amplitude of the SK-mediated I_{AHP} in infected rat neurons. Thus, by increasing or reducing the amount of rSK1 channels expressed in a heterologous system or rat neurons, we can decrease or increase rSK2 channels expressed on the cell membrane.

Previous studies have shown that when expressed in HEK293 cells, the interaction of rSK1 and rSK2 channels resulted in an overall larger SK-channels mediated current and a change in their pharmacology (Benton et al., 2003; Church et al., 2015). This appears to be in contrast with our results in both Cosm6 cells and rat neurons where overexpression of rSK1-HA reduced the amount of rSK2 channels expressed in the cell membrane. However, it remains possible that these channels assemble as heteromultimers with the total fraction of SK1-SK2 heteromeric protein in the membrane being very low.

Our results suggest that in the rodent brain rSK1 channels, rather than acting as independent membrane ion channels, are involved in trafficking of rSK2 channels to the plasma membrane. hSK1, which is ~84% homologous to the rodent channels (D'Hoedt et al., 2004), behaves entirely differently, being translated as a functional ion channel (Church et al., 2015). The reason for this difference is not clear. It is possible, that the rSK1-rSK2 interaction and the regulation of rSK2 channels expressed on the cell membrane is important during development. rSK1 and rSK2 channel transcripts within the hippocampus (CA1, CA3 and dentate gyrus) start being expressed at embryonic day 19 (E19) in rodents (Gymnopoulos et al., 2014), with their expression patterns being very similar, showing colocalization (Stocker and Pedarzani, 2000). These similar expression patterns could indicate a developmental change within the expression of rSK2 channels on the cell membrane that could affect neuronal physiology. Interestingly, rSK2 was found to be present mainly in the ER of CA1 pyramidal neurons at P5. By P30, however, most rSK2 was present at spines and dendrites changing the cellular physiology (Ballesteros-Merino et al., 2012). It is, therefore, possible that rSK1 channels modulate the amount of rSK2 channels expressed on the cell membrane. While rSK2 channels are trapped in the ER by rSK1 channels, neurons can receive more inputs and increased the excitation in the early developmental stages, which will increase the memory acquisition. Once later developmental stages are reached, however, rSK1 channels may have a different role in regulating the expression of rSK2 channels and the neuronal physiology.

ETHICS STATEMENT

This was approved by the University of QLD Ethics Committee.

AUTHOR CONTRIBUTIONS

EA discussed experiments, did experiments, and wrote the manuscript. PSe did experiments and wrote the manuscript. LX made virus. MR did experiments and made figures. AT did experiments, made figures and wrote the manuscript. PSa designed experiments, made figures and wrote the manuscript.

FUNDING

ACKNOWLEDGMENTS

REFERENCES

Adelman, J. P., Maylie, J., and Sah, P. (2012). Small-conductance Ca^{2+}-activated K$^+$ channels: form and function. *Ann. Rev. Physiol.* 74, 245–269. doi: 10.1146/annurev-physiol-020911-153336

Ballesteros-Merino, C., Lin, M., Wu, W. W., Ferrandiz-Huertas, C., Cabanero, M. J., Watanabe, M., et al. (2012). Developmental profile of SK2 channel expression and function in CA1 neurons. *Hippocampus* 22, 1467–1480. doi: 10.1002/hipo.20986

Benton, D. C., Monaghan, A. S., Hosseini, R., Bahia, P. K., Haylett, D. G., and Moss, G. W. (2003). Small conductance Ca^{2+}-activated K$^+$ channels formed by the expression of rat SK1 and SK2 genes in HEK 293 cells. *J. Physiol.* 553, 13–19. doi: 10.1113/jphysiol.2003.054551

Bond, C. T., Herson, P. S., Strassmaier, T., Hammond, R., Stackman, R., Maylie, J., et al. (2004). Small conductance Ca^{2+}-activated K$^+$ channel knock-out mice reveal the identity of calcium-dependent afterhyperpolarization currents. *J. Neurosci.* 24, 5301–5306. doi: 10.1523/jneurosci.0182-04.2004

Church, T. W., Weatherall, K. L., Correa, S. A., Prole, D. L., Brown, J. T., and Marrion, N. V. (2015). Preferential assembly of heteromeric small conductance calcium-activated potassium channels. *Eur. J. Neurosci.* 41, 305–315. doi: 10.1111/ejn.12789

Coetzee, W. A., Amarillo, Y., Chiu, J., Chow, A., Lau, D., McCormack, T., et al. (1999). Molecular diversity of K$^+$ channels. *Ann. N Y Acad. Sci.* 868, 233–285. doi: 10.1111/j.1749-6632.1999.tb11293.x

Delaney, A. J., Sedlak, P. L., Autuori, E., Power, J. M., and Sah, P. (2013). Synaptic NMDA receptors in basolateral amygdala principal neurons are triheteromeric proteins: physiological role of GluN2B subunits. *J. Neurophysiol.* 109, 1391–1402. doi: 10.1152/jn.00176.2012

Desai, R., Peretz, A., Idelson, H., Lazarovici, P., and Attali, B. (2000). Ca^{2+}-activated K$^+$ channels in human leukemic Jurkat T cells. Molecular cloning, biochemical and functional characterization. *J. Biol. Chem.* 275, 39954–39963. doi: 10.1074/jbc.m001562200

D'Hoedt, D., Hirzel, K., Pedarzani, P., and Stocker, M. (2004). Domain analysis of the calcium-activated potassium channel SK1 from rat brain. Functional expression and toxin sensitivity. *J. Biol. Chem.* 279, 12088–12092. doi: 10.1074/jbc.c300382200

Faber, E. S., Delaney, A. J., and Sah, P. (2005). SK channels regulate excitatory synaptic transmission and plasticity in the lateral amygdala. *Nat. Neurosci.* 8, 635–641. doi: 10.1038/nn1450

Fiorillo, C. D., and Williams, J. T. (1998). Glutamate mediates an inhibitory postsynaptic potential in dopamine neurons. *Nature* 394, 78–82. doi: 10.1038/27919

Gymnopoulos, M., Cingolani, L. A., Pedarzani, P., and Stocker, M. (2014). Developmental mapping of small-conductance calcium-activated potassium channel expression in the rat nervous system. *J. Comp. Neurol.* 522, 1072–1101. doi: 10.1002/cne.23466

Ishii, T. M., Maylie, J., and Adelman, J. P. (1997). Determinants of apamin and d-tubocurarine block in SK potassium channels. *J. Biol. Chem.* 272, 23195–23200. doi: 10.1074/jbc.272.37.23195

Joiner, W. J., Wang, L.-Y., Tang, M. D., and Kaczmarek, L. K. (1997). hSK4, a member of a novel subfamily of calcium-activated potassium channels. *Proc. Natl. Acad. Sci. U S A* 94, 11013–11018. doi: 10.1073/pnas.94.20.11013

Köhler, M., Hischberg, B., Bond, C. T., Kinzie, J. M., Marrion, N. V., Maylie, J., et al. (1996). Small-conductance, calcium-activated potassium channels from mammalian brain. *Science* 273, 1709–1714. doi: 10.1126/science.273.5282.1709

Lee, C. H., and MacKinnon, R. (2018). Activation mechanism of a human SK-calmodulin channel complex elucidated by cryo-EM structures. *Science* 360, 508–513. doi: 10.1126/science.aas9466

Lin, M. T., Luján, R., Watanabe, M., Adelman, J. P., and Maylie, J. (2008). SK2 channel plasticity contributes to LTP at Schaffer collateral-CA1 synapses. *Nat. Neurosci.* 11, 170–177. doi: 10.1038/nn2041

Ngo-Anh, T. J., Bloodgood, B. L., Lin, M., Sabatini, B. L., Maylie, J., and Adelman, J. P. (2005). SK channels and NMDA receptors form a Ca^{2+}-mediated feedback loop in dendritic spines. *Nat. Neurosci.* 8, 642–649. doi: 10.1038/nn1449

Sah, P. (1996). Ca^{2+}-activated K$^+$ currents in neurones: types, physiological roles and modulation. *Trends Neurosci.* 19, 150–154. doi: 10.1016/s0166-2236(96)80026-9

Sailer, C. A., Hu, H., Kaufmann, W. A., Trieb, M., Schwarzer, C., Storm, J. F., et al. (2002). Regional differences in distribution and functional expression of small-conductance Ca^{2+}-activated K$^+$ channels in rat brain. *J. Neurosci.* 22, 9698–9707. doi: 10.1523/jneurosci.22-22-09698.2002

Stocker, M., Hirzel, K., D'Hoedt, D., and Pedarzani, P. (2004). Matching molecules to function: neuronal Ca^{2+}-activated K$^+$ channels and afterhyperpolarizations. *Toxicon* 43, 933–949. doi: 10.1016/j.toxicon.2003.12.009

Stocker, M., Krause, M., and Pedarzani, P. (1999). An apamin-sensitive Ca^{2+}-activated K$^+$ current in hippocampal pyramidal neurons. *Proc. Natl. Acad. Sci. U S A* 96, 4662–4667. doi: 10.1073/pnas.96.8.4662

Stocker, M., and Pedarzani, P. (2000). Differential distributions of three Ca^{2+}-activated K$^+$ channel subunits, SK1, Sk2, and SK3, in the adult rat central nervous system. *Mol. Cell. Neurosci.* 15, 476–493. doi: 10.1006/mcne.2000.0842

Strassmaier, T., Bond, C. T., Sailer, C. A., Knaus, H. G., Maylie, J., and Adelman, J. P. (2005). A novel isoform of SK2 assembles with other SK subunits in mouse brain. *J. Biol. Chem.* 280, 21231–21236. doi: 10.1074/jbc.m413125200

Vergara, C., Latorre, R., Marrion, N. V., and Adelman, J. P. (1998). Calcium-activated potassium channels. *Curr. Opin. Neurobiol.* 8, 321–329. doi: 10.1016/S0959-4388(98)80056-1

Xia, X. M., Fakler, B., Rivard, A., Wayman, G., Johnson-Pais, T., Keen, J. E., et al. (1998). Mechanism of calcium gating in small-conductance calcium-activated potassium channels. *Nature* 395, 503–507. doi: 10.1038/26758

SUPPLEMENTARY MATERIAL

Hierarchical and Nonlinear Dynamics in Prefrontal Cortex Regulate the Precision of Perceptual Beliefs

Leonardo L. Gollo [1,2] *[†], Muhsin Karim [3,4][†], Justin A. Harris [5], John W. Morley [6] and Michael Breakspear [1,2,3,4,7,8] *

[1]QIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia, [2]Centre of Excellence for Integrative Brain Function, QIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia, [3]School of Psychiatry, Faculty of Medicine, University of New South Wales, Sydney, NSW, Australia, [4]The Black Dog Institute, Sydney, NSW, Australia, [5]School of Psychology, The University of Sydney, Sydney, NSW, Australia, [6]School of Medicine, Western Sydney University, Sydney, NSW, Australia, [7]Metro North Mental Health Service, Brisbane, QLD, Australia, [8]Hunter Medical Research Institute, University of Newcastle, New Lambton Heights, NSW, Australia

Actions are shaped not only by the content of our percepts but also by our confidence in them. To study the cortical representation of perceptual precision in decision making, we acquired functional imaging data whilst participants performed two vibrotactile forced-choice discrimination tasks: a fast-slow judgment, and a same-different judgment. The first task requires a comparison of the perceived vibrotactile frequencies to decide which one is faster. However, the second task requires that the estimated difference between those frequencies is weighed against the precision of each percept—if both stimuli are very precisely perceived, then any slight difference is more likely to be identified than if the percepts are uncertain. We additionally presented either pure sinusoidal or temporally degraded "noisy" stimuli, whose frequency/period differed slightly from cycle to cycle. In this way, we were able to manipulate the perceptual precision. We report a constellation of cortical regions in the rostral prefrontal cortex (PFC), dorsolateral PFC (DLPFC) and superior frontal gyrus (SFG) associated with the perception of stimulus difference, the presence of stimulus noise and the interaction between these factors. Dynamic causal modeling (DCM) of these data suggested a nonlinear, hierarchical model, whereby activity in the rostral PFC (evoked by the presence of stimulus noise) mutually interacts with activity in the DLPFC (evoked by stimulus differences). This model of effective connectivity outperformed competing models with serial and parallel interactions, hence providing a unique insight into the hierarchical architecture underlying the representation and appraisal of perceptual belief and precision in the PFC.

Keywords: decision making, dynamic causal modeling, fMRI, prefrontal cortex, vibrotactile

INTRODUCTION

Percepts underpin all our interactions with the world. Perceptual precision, the confidence with which we hold those percepts, informs this interaction, such as when a decision is biased toward a precisely represented percept (Ernst and Banks, 2002). Although high perceptual precision may be advantageous in some contexts, such as when driving a car, there exist other situations where a degree of imprecision is crucial: if percepts were held with infinite

precision then it would be impossible to recognize any object encountered for a second time. For example, the texture of a surface would feel unique and surprising on every touch. Whereas the neurobiology of perception has been a long-studied subject, research into the basis of perceptual precision and its impact on decision making has been more recent (Knill and Pouget, 2004; Moran et al., 2013; Pouget et al., 2013; Navajas et al., 2017).

The neural basis of perceptual decision-making has been extensively studied using two-alternative forced-choice tasks in the somatosensory (Romo and Salinas, 2003) and visual domain (Britten et al., 1992). These prototypical experiments consist in presenting two sequential stimuli that are followed by a forced response between two choices involving a comparison between the properties of these two stimuli (see **Figure 1**). In the somatosensory modality, a wealth of neurophysiological research using vibrotactile stimuli has established the crucial role of the prefrontal cortex (PFC) during the performance of such tasks (Gold and Shadlen, 2007; Hegner et al., 2007; Heekeren et al., 2008; Wang, 2012). While the primary somatosensory cortex is clearly involved in stimulus representation (Hernández et al., 2000; Harris et al., 2002; Sörös et al., 2007), the PFC holds the representation of the first stimulus in working memory for subsequent comparison against representation of the second stimulus (Preuschhof et al., 2006; Wang, 2008), as well as the final decision process (Miller et al., 2003; Pleger et al., 2006; Heekeren et al., 2008; Wang, 2008; Barak et al., 2010). With very few exceptions (Engel and Wang, 2011), decisions in these forced-choice experiments are only dependent on magnitude comparisons of the perceived frequencies. A sensory percept can be viewed probabilistically (as a probability distribution) and to first order can hence be decomposed into its magnitude (here, the perceived frequency) and its precision (the inverse of the variance of the probability distribution; see **Figure 2**). Whilst perceptual precision—classically captured by the signal-to-noise ratio—impacts upon the performance accuracy of a faster-slower comparison, the decision itself does not explicitly require representing and acting on the precision of those perceptions. This is because the final decision only rests upon deciding whether the second stimulus is faster or slower than the first and does not depend upon the subjective confidence in that judgment. That is, a faster-slower decision can be made by a simple subtraction and does not crucially depend upon the precision of either percept.

The anterior cingulate and ventromedial PFC appear to play critical roles in assessing the value of current information in an environment of uncertain outcome and reward (Daw et al., 2005; Kennerley et al., 2006; Behrens et al., 2007). These regions also represent changes in this value (that is, when the link between stimulus, outcome and reward is volatile; Rushworth and Behrens, 2008). Whilst the value of the percept to an external reward is uncertain in these studies (Fiorillo et al., 2003; Yu and Dayan, 2005; Hsu et al., 2005; Huettel et al., 2006; Behrens et al., 2007; Tobler et al., 2007), the percept itself is not ambiguous. Hence, it is not clear from these studies whether these regions are also involved

in representing the intrinsic precision of the percept itself, or whether other regions are recruited when the stimulus is noisy but the task contingencies are fixed (Kayser et al., 2010; Bach and Dolan, 2012).

Here, we sought to disentangle the representation of stimulus properties from the precision of those representations in the PFC. Functional neuroimaging data were acquired whilst paired vibrotactile flutter stimuli (10–50 Hz) were sequentially applied to the index finger. In separate tasks, participants were requested to decide if the second stimulus was *faster* than the first, or if the second stimulus was *different* from the first. As rehearsed above, the "faster-slower" task can be performed by simply encoding and subtracting an estimate of each stimulus frequency—that is, decisions only explicitly depend on comparing the likely value of each of the flutter frequencies. In the "same-different" task, the magnitude of this subtraction must be weighed against the precision of the perceptual beliefs, such that a difference that is perceived as small may be inferred as significant if each percept is held precisely (and conversely for imprecise representations). The precision of a percept is the composite of the roughness of the stimulus and the perceptual imprecision due to stochastic effects in perceptual systems: to manipulate stimulus precision, noise was introduced to the vibrotactile oscillatory frequency as an additional experimental factor (Harris, 2006; Harris et al., 2006; Karim et al., 2012). Note that we refer to precision in the statistical sense of the inverse of the noise variance (**Figure 2**).

The PFC is known to be underpinned by extensive intrinsic anatomical connections, forming local circuits that adapt to contextual demands at hand (Fuster, 2001; Miller and Cohen, 2001; Botvinick, 2008). The hierarchical nature of these circuits during the representation of perceptual precision is poorly understood (Nee and D'Esposito, 2016). We first identify a constellation of regions in the left PFC that respond to these stimulus and task manipulations. We then study the prefrontal networks that underpin our data using dynamic causal modeling (DCM). DCM is a model-based technique to infer network dynamics (Friston et al., 2003) that has found explanatory utility in cognitive neuroscience, including language (Leff et al., 2008; Noppeney et al., 2008), motor processes (Grefkes et al., 2008), vision (Mechelli et al., 2003; Fairhall and Ishai, 2007) and memory (Smith et al., 2006). DCM has been employed to study perceptual decision-making tasks (Summerfield et al., 2006; Stephan et al., 2007; Summerfield and Koechlin, 2008) including vibrotactile discrimination tasks, focussing on the exchange of information from primary to secondary somatosensory cortex (Kalberlah et al., 2013). Here, we use DCM to disambiguate between candidate serial, parallel or hierarchical engagement of the PFC in the representation and manipulation of perceptual precision.

MATERIALS AND METHODS

Overview

Sixteen healthy young adults participated in our experiment. To avoid ceiling or floor effects and reduce inter-subject variability in performance, participants first performed an adaptive staircase
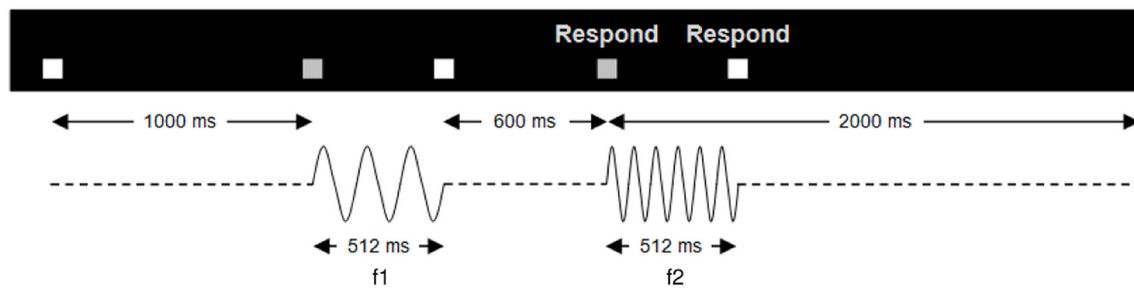
FIGURE 1 | Trial structure. Temporal structure of a single trial of the vibrotactile discrimination task. A pair of stimuli (f1, f2), each 512 ms in duration, separated by an ISI of 600 ms, was presented to the participant's right index finger. The start of the trial was indicated by a white box, which turned gray when the vibrations were presented. Upon the onset of the second vibration, a respond screen appeared indicating that the participant could make a button press. Participants had 2 s in which to respond after the second vibration onset. Trials were presented in four sessions; two sessions of faster-slower and two same-different.
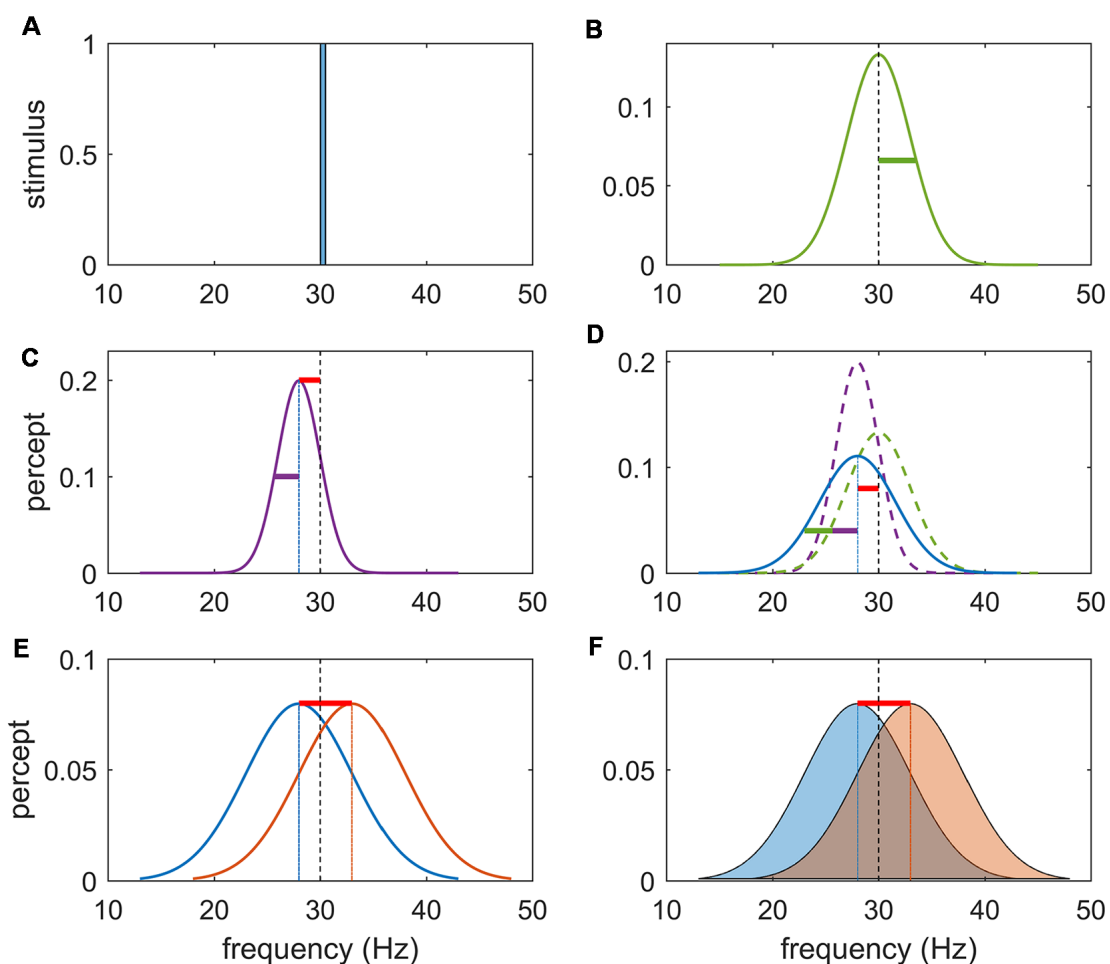
FIGURE 2 | Schema for task rationale. **(A)** Frequency content of a noise-free stimulus of 30 Hz. **(B)** Noise imbued vibrotactile stimulus with center frequency of 30 Hz and variance of stimulus noise represented by the green bar. Precision refers to the inverse of the variance of the percept. **(C)** Perceptual encoding of a noise-free stimulus can be represented by a unimodal distribution centered at the likely value of the inferred stimuli. Note that due to an inevitable perceptual error (bias) this inferred stimulus is shifted to the left (or right) of the true stimulus frequency (red bar) and has perceptual noise (purple bar). **(D)** Perceptual representation of a noisy stimulus can be conceptualized as the sum of the stimulus (external) noise (green) and the perceptual (internal) noise (purple). It may have a perceptual bias (red bar) and perceptual noise (purple bar) in addition to stimulus noise (green bar). In separate sessions, participants were either instructed to answer the question "Is the 2nd vibration faster?" or "Are the vibrations different?" as a yes/no response. **(E)** The first task can be solved by subtracting the values of the inferred stimulus and responding on the sign of the answer. **(F)** The second task requires that the inferred magnitude of this difference be weighted by the precision (inverse variance) of each percept. Due to the perceptual error, there will exist a difference in the inferred frequency difference even if f1 = f2.

procedure. Behavioral and functional imaging data were then acquired while they performed the main vibrotactile experiment. Analyses of these data then informed the employment of DCM. Each of these steps is described below. Full details are provided in the **Supplementary Material**.

Participants

Sixteen healthy volunteers (10 men; mean age, 28.4 years; standard deviation, 9.3; age range, 20–61 years) participated in the study. Participants gave written informed consent and the study was approved by the University of New South Wales Human Research Ethics Committee. Participants were paid for their participation in the study. All participants were right-handed. Participants disavowed history of a psychiatric disorder, neurological disorder, or drug or alcohol dependence. Participants gave written informed consent according to local institutional human ethics committee approval.

Stimuli and Task

Using an MR-compatible stimulator, mechanical vibrotactile stimuli were delivered to the right index finger (see **Supplementary Material, SM1.1**). Trials consisted of a series of paired stimuli, each 512 ms in duration, separated by an ISI of 600 ms (**Figure 1** and **Supplementary Material, SM1.2**).

Titration Procedure

To limit individual variability in performance and avoid ceiling effects in accuracy, we used a titration procedure that matched average task performance *via* an adaptive staircase procedure as described previously (Karim et al., 2012). The participants responded to the question: "Is the 2nd vibration faster?" For each trial, one of the vibrations was the base 34 Hz, and the other a comparison vibration, which varied based on the participant's current performance according to an adaptive staircase procedure. The presentation order of the base and comparison was pseudorandomly varied from trial to trial.

Two intermixed staircases (easy and hard) selected at random were used to limit the participant from experiencing a learning effect from consecutive easy or consecutive hard trials. The difference in frequency between vibration pairs was initially set to 5 Hz, then progressively decreased or increased by 10% of the current frequency difference. For both staircases, a step-up occurred for each incorrect response. For the easy staircase, a step-down occurred after six non-consecutive correct responses. That is, even amongst trials of incorrect responses, a tally was kept for each correct response made. Once the tally reached six, a step-down occurred and the tally was reset to zero. Likewise, for the hard staircase, a step-down occurred after two non-consecutive correct responses. We sought to have performance converge at ∼90% and ∼65% proportion correct, respectively (Zwislocki and Relkin, 2001). A medium value of difficulty (target accuracy of 75%) was determined by calculating the geometric mean between the easy and hard frequency differences (Karim et al., 2012).

Behavioral Task

Following titration, participants completed a parametric vibrotactile discrimination task with factors of context, noise

and difficulty. "Context" denotes the task instructions—the faster/slower or the same/different comparison; "noise" refers to the presence or absence of random fluctuations in the stimuli. "Difficulty" refers to the (titrated) difference between the stimulus frequencies.

To create the noise factor, the temporal structure of the two vibrations was degraded by adding independent Gaussian-distributed values (mean = 0) to the wavelength of each cycle of the sine wave (Harris et al., 2006). We added 8% noise so that the standard deviation of the cycle length within the vibration equalled 0.08 of the base cycle length. For example, a 40 Hz vibration was comprised of cycles with mean length 25 ms and standard deviation of 2 ms. We hence refer to all trials as "regular" (noise-free) or "noisy."

The contextual (task) factor was created by asking participants to perform either a fast-slow or a same-different comparison. In the fast-slow task, participants were instructed to answer the question "Is the 2nd vibration faster?" as a yes/no response. They were informed that there was always a faster vibration (i.e., no identical trials). In the same-different task, participants were instructed to answer the question "Are the vibrations different?" as a yes/no response. They were (correctly) informed that half of the presented vibration pairs were the same and half were different. Different trials in the second (same/different) context were identical to the corresponding trials in the first (faster/slower) context. For same-noisy trials in the second context, exactly the same stimulus was presented—that is, both the center frequency and the exact same pseudorandom sequence of jittered wavelengths. The rationale for our task design is illustrated in **Figure 2**.

For feasibility issues, not all cells in the full factorial design were performed. For example, in pilot testing, the accuracy of hard-noisy trials was at chance (50%) and was thus not used. We refer to the task as a "partial" factorial design in this sense. We do not report on the effect of task difficulty in this article and hence collapse all available trials (of equivalent difficulty) across this factor (for further details, see **Supplementary Material, SM1.3** and **Supplementary Table S1**).

MRI Acquisition and Analysis

Functional imaging data were acquired using a Philips (Achieva X) 3.0-Tesla scanner (for acquisition details see **Supplementary Material, SM1.4**). Stimuli were delivered *via* the vibrotactile device to the right index finger. Participants made button press responses *via* their left index and middle fingers. Inter-trial intervals were pseudorandomly jittered between 6 and 12 s to decorrelate the evoked hemodynamic responses between trials. The task was conducted over four separate sessions separated by a short break. Each block consisted of exclusively same-different or faster-slower trials. Pre-processing of dynamic images included realignment, normalization, re-sampling and spatial smoothing using SPM8. Statistical analysis of the time series of images was conducted using the General Linear Model (GLM; Friston et al., 1994a) with regressors modeling each of the factor components. To focus on the decision-making process, we used a boxcar of width 200 ms immediately prior to the button press response. The results reported here are robust to changes

in the width of the regressor. These were convolved with the canonical hemeodynamic response function.

Group-level, random-effects analyses used a flexible factorial analysis of variance (ANOVA) including a subject factor and non-sphericity correction for repeated measures (i.e., inhomogeneity of variance among conditions was estimated with ReML). In the second (same-different) task there also exists an additional stimulus factor, namely "Different" vs. "Same" trials: we hence also investigate this factor within this context. Statistical inference was performed at the cluster-level using family-wise error (FWE) correction, $p < 0.05$ (Friston et al., 1994b, 1996). Unless otherwise stated, we employed a height threshold of $p < 0.00005$ and a spatial extent of 20 voxels. All p-values reported in the Results are FWE-corrected. Cluster locations were identified using the SPM Anatomy toolbox (Eickhoff et al., 2005).

Dynamic Causal Modeling

Model Specification
DCM is a computational approach that allows construction and comparison of dynamic network models of functional imaging data (Friston et al., 2003). DCM uses the time series from imaging data and combines a model of the hidden neuronal dynamics with a forward model that translates neural states into predicted measurements (Stephan et al., 2008). Specifying dynamic causal models requires two steps: first, regions (network "nodes") that express the specific effects of interest (noise, context, same-different) are identified using the preceding GLM. These are described in the "Results" section, following analysis of the main and the interaction effects in our experiment. The time series data from each node are then extracted. We used a sphere of 6 mm radius centered at the voxel showing the group-wise maximum contrast (see **Supplementary Material, SM1.5.1**).

The second step in DCM specification involves the construction of a space of models that embody various hypotheses about the manner in which these nodes interact—that is, the (effective) connectivity, or network "edges," between the nodes. Restricting the space of models to a relatively small family that test specific hypotheses is an important way to constrain the number (and utility) of models to be tested (Stephan et al., 2010). Since the present objective was to use DCM to study the network models of perceptual precision (hence, not focussing on basic vibrotactile processing *per se*), we restricted our analyses to a small number of models that shared a common sensory input base and added candidate integrative mechanisms on top of this base. The input base was the sensory area showing the main effect of stimuli, hence identified using an F-contrast across all trials. We introduced eight separate models (four bilinear and three nonlinear) on top of the common base that modeled serial or parallel integrative mechanisms. Serial, parallel or hierarchical architectures play varying roles in a diversity of cognitive and even machine learning systems (Mesulam, 1998; Friston, 2005; Petersen and Sporns, 2015): their disambiguation here, using DCM, can hence contribute to this broader literature, whilst also establishing the relative primacy of perceptual value vs. precision underlying decision-making in the presence of stimulus noise. These

DCM's each embody one of these arrangements, differing within-class according to the presence or absence of symmetrical relationships (see **Figure 5**, and results for a representation of the specified models). Nonlinear models specify hierarchical relationships between the network nodes—that is, where the neuronal activity in one region gates the flow of activity between other regions (Stephan et al., 2010); bilinear models mirror their more complex nonlinear counterparts, except they lack hierarchical relationships between regions: this gating (interaction) function is instead fulfilled by non-specific modulatory inputs.

Model Selection and Parameter Estimation
Following model specification, DCM employs Bayesian model selection (BMS) to identify which model is the most likely to have generated the observed data. The process of adjudicating between models essentially balances their goodness of fit against a factor that penalises models for their relative complexity (for review, see Marreiros et al., 2010). BMS yields the evidence for each model—the (posterior) probability of the model given the data—as well as the estimated (posterior) parameter values that reflect the strength of interactions between regions. Relative evidence for all models is used to identify the most likely model, or the best family of models (see **Supplementary Material, SM1.5.2**). We performed BMS using random effects analysis (Stephan et al., 2009).

RESULTS

Behavioral Results
Analysis of the behavioral data revealed significant effects of both context and noise (**Table 1**, **Figure 2**; also **Supplementary Material, SM2.1** and **Supplementary Figure S1**): consistent with its lesser computational burden, participants were more accurate and had faster response times (RTs) for the fast-slow task compared to the same-different one (see **Figure 3A**, and for effect sizes, see **Table 1**[1a,1b]). Across both contexts, there was also a significant effect of noise: the presence of aperiodic temporal noise in the vibrotactile stimuli decreased accuracy[1c] across both contexts and slowed RT for the same-different context[1d]. There was no significant interaction between context and noise.

The lower accuracy in the same-different compared to the faster-slower context could in theory be due to a response bias arising, for example, from a conservative internal standard for the detection of difference. We estimated d-prime (d'), a measure of sensitivity that takes response bias into account (MacMillan and Creelman, 2005). Repeated measures ANOVA re-affirmed significantly lower accuracy for responses in the same-different compared to the fast-slow context (d' for fast-slow = 1.59, d' for same-different = 0.72, $F_{(1,15)} = 36.497$, $p < 0.0001$). This suggests that differences in the same-different context were associated with a loss in sensitivity.

Within the same-different task, participants took longer to respond to the same compared to the different trials (**Figure 3A**)[1e,f] which was associated with a trend-level increase in accuracy ($p = 0.0509$). There was an interesting interaction between noise and difference for accuracy[1g]: for

TABLE 1 | Behavioral performance statistics for context (Fast-slow, Same-different), noise (Regular, Noisy), and difference (Different, Same).

Contrast	Factor	Dependent variable	F-statistic	p-value	Partial eta square	Text ref.
Noise and context	Context	PC	$F_{(1,15)} = 87.039$	$p < 0.0001*$	0.853	1a
	Noise		$F_{(1,15)} = 5.352$	$p = 0.0353*$	0.263	1c
	Noise * Context		$F_{(1,15)} = 0.672$	$p = 0.4251$	0.043	
(Both contexts)	Context	RT	$F_{(1,15)} = 28.759$	$p = 0.0001*$	0.657	1b
	Noise		$F_{(1,15)} = 3.154$	$p = 0.0960$	0.174	
	Noise * Context		$F_{(1,15)} = 0.419$	$p = 0.5273$	0.027	
Noise and difference	Noise	PC	$F_{(1,15)} = 0.009$	$p = 0.9256$	0.001	1h
	Difference		$F_{(1,15)} = 4.502$	$p = 0.0509$	0.231	1f
	Noise * Difference		$F_{(1,15)} = 17.927$	$p = 0.0007*$	0.544	1g
(Same-different context only)	Noise	RT	$F_{(1,15)} = 7.240$	$p = 0.0168*$	0.326	1d
	Difference		$F_{(1,15)} = 19.225$	$p = 0.0005*$	0.562	1e
	Noise * Difference		$F_{(1,15)} = 0.286$	$p = 0.6008$	0.019	

*Proportion correct (PC) was used to assess accuracy and response time (RT) was used to assess speed. *Significant p-values.*

FIGURE 3 | Behavioral results for the same-different context and interpretations. **(A)** Reaction time for Fast-Slow vs. Same-Different comparisons. Note the longer reaction times for the latter task. **(B)** Proportion of correct (PC) responses (or accuracy) of regular and noisy response for different and same trials in the Same-Different task. **(C)** Stimulus noise increases the variance of the perceptual representation of the two frequencies f1 and f2, increasing the overlap between them. A larger overlap between perceptual representations decreases the sensitivity of responses to Different trials (left). The yellow bar depicts the difference between the mean of the two percepts—here the sum of the true stimulus differences and the perceptual error. Conversely, noise increases the accuracy of responses to Same trials (right): some slight difference in perception occurs even for identical, periodic stimuli (red bars, sum of perceptual errors). However, stimulus noise degrades the precision of each percept, hence increasing their overlap and masking these small (false) perceptual differences.

same trials, accuracy was greatest when trials were noisy, whereas for different trials accuracy was higher for regular trials (**Figure 3B**, $p < 0.0007$).

Thus, it appears easier for participants to correctly classify same trials as "same" when they are imbued with temporal noise than when they are pure sinusoids. Conversely, different trials were more likely to be correctly reported when they are regular. These observations can be interpreted by considering the influence of stimulus noise on perceptual accuracy (**Figure 3**): we return to this issue in the "Discussion" section.

Functional Imaging Contrasts

We observed a strong and significant main effect of "context" in our functional imaging data, with several clusters surviving FWE-corrected significance (**Table 2** and **Supplementary Material, SM2.2.1**). All of these effects were in the direction of the same-different over the fast-slow context, again consistent with the additional computational load of this task and mirroring the behavioral results. The strongest effect was expressed in a large cluster in the left inferior frontal gyrus (BA 45; $p < 0.0001$, **Supplementary Figure S2A**), occupying the

mid-ventrolateral PFC (VLPFC). A second effect was observed in the right middle temporal gyrus (BA 21; $p < 0.0001$, **Supplementary Figure S2B**). Also in accordance with the behavioral results, no significant interaction effects between noise and context were found.

We next focussed on effects present within the same-different context (**Table 3**, **Supplementary Material, SM2.2.2**). The contrast of different over same trials yielded three distinct clusters, all of which surpassed FWE-corrected significance for both cluster and height statistics. The strongest effect was centered over the left inferior parietal lobule (BA 40; $p < 0.0001$, **Figure 4A**) and included voxels within the supramarginal and the post-central gyri. Other effects occurred in the PFC, including a strong effect in the left middle frontal gyrus (the dorsolateral PFC, DLPFC, BA 44; $p < 0.002$, **Figure 4B**). Inspection of the parameter values for these two regions revealed quite distinct responses: whereas the large posterior cluster showed significantly positive values for both different and same trials (with the different greater than same trials, consistent with repetition suppression), the DLPFC cluster only showed non-zero responses to different trials, specific to the "signal trials" (true positives) in this context. A third cluster was located in the midline, centered on the supplementary motor area (BA 6; $p < 0.008$).

The contrast between regular and noisy trials speaks directly to the representation of perceptual precision. Interestingly, despite the absence of a significant effect of stimulus noise on behavioral accuracy in same-different trials[1h], there existed a strong and specific effect in the imaging data, with a single cluster towards the rostral pole of the left PFC, and in the left DLPFC, for the contrast of regular over noisy trials (BA 10; $p < 0.016$, FWE-corrected, **Figure 4C**). This cluster lies within a sulcus in rostral PFC (rPFC, BA10), bounded dorsally by the DLPFC. There were no effects approaching significance for the contrast of noisy over regular trials.

The significant interaction between regular-noisy trials and same-different trials present in the behavioral data[1g] motivated analysis of the corresponding interaction in the functional imaging data. We observe a single significant cluster, located within the left superior frontal gyrus (SFG, BA 8, $p < 0.010$ FWE-corrected, **Figure 4D**, **Table 3**).

We, therefore, observe four distinct clusters in the left PFC for the main effect of context, the main effect of noise, the main effect of difference and the interaction between noise and difference. Whilst nearby, these four clusters nonetheless reside in distinct sulci. One cluster resides with the VLPFC, and two within the DLPFC.

Dynamic Causal Modeling

We next employed DCM to model the interactions between the left inferior parietal lobe (IPL) and the three prefrontal clusters engaged in the second (faster-slower) context (**Figure 4E** and **Supplementary Material, SM3.5**). We excluded areas outside of the PFC, such as the supplementary motor area, likely involved in lower level processing and/or preparation for the motor response. All specified dynamic causal models of these data shared a common input base, beginning with stimulus inputs (i.e., vibrotactile stimuli) directed to the left IPL. The effect of regular trials expressed in the rPFC was modeled by an effective connection from IPL to rPFC, modulated by the pure (regular) trials (**Figure 4E**). Likewise, a connection from the IPL to the DLPFC, modulated by stimulus difference, modeled the effect of difference observed in the DLPFC. Finally, SFG is subjected to the influence of both modulations as the interaction between regular-noisy and same-different trials occurs there. Note that a backward connection was placed here to allow for the diminished response of different compared to same trials to be modeled by the feedback influence of the DLPFC on the IPL.

We specified seven separate models (four bilinear: "Diamond," "Fork," "Legs 1," and "Legs 2"; and three nonlinear: "Stork 1," "Stork 2," and "Stork 3"; see **Supplementary Material** for additional details) on top of this common base that represent serial, parallel or hierarchical processes (see "Materials and Methods" section and **Figure 5**). As the name suggests, in serial models (both bilinear and nonlinear), information passes in a serial manner from the IPL *via* the rPFC or the DLPFC (or both)

TABLE 2 | Significant clusters for the effect of context (fast-slow vs. same-different).

Contrast	Anatomical label	R/L	MNI coordinates			BA	T-value	Statistics	
			X	Y	Z			Cluster P$_{FWE-corr}$	Size (voxels)
T: Fast-slow < Same-different	Inferior Frontal Gyrus pars triangularis	L	−36	29	16	45	6.43	<0.0001	41
	Middle Temporal Gyrus	R	51	−25	−14	21	5.59	<0.0001	36

Significant results of "Faster-slower" < "Same-different" are shown. Standard Montreal Neurological Institute (MNI) coordinates correspond to peak maxima. Size indicates the number of voxels in the cluster. Note that the "Same" trials have been omitted from the same-different contrast as there were no counterpart same trials from the fast-slow contrast.

TABLE 3 | Significant clusters for contrasts within the same-different context.

| Contrast	Anatomical label anatomy	R/L	MNI coordinates			BA	T-stat	Statistics	
			X	Y	Z			Cluster P$_{FWE-corr}$	Extent
Different > Same	Inferior Parietal Lobule	L	−45	−46	43	40	5.81	0.0001	145
	Middle Frontal Gyrus (DLPFC)	L	−39	14	34	44	5.69	0.002	42
	Supplementary motor area	L	0	23	52	6	4.91	0.008	29
Regular > Noisy	Middle Frontal Gyrus (PFC)	L	−27	44	19	10	4.78	0.016	22
Noise × Difference	Superior Frontal Gyrus	L	−21	23	37	8	4.88	0.010	26
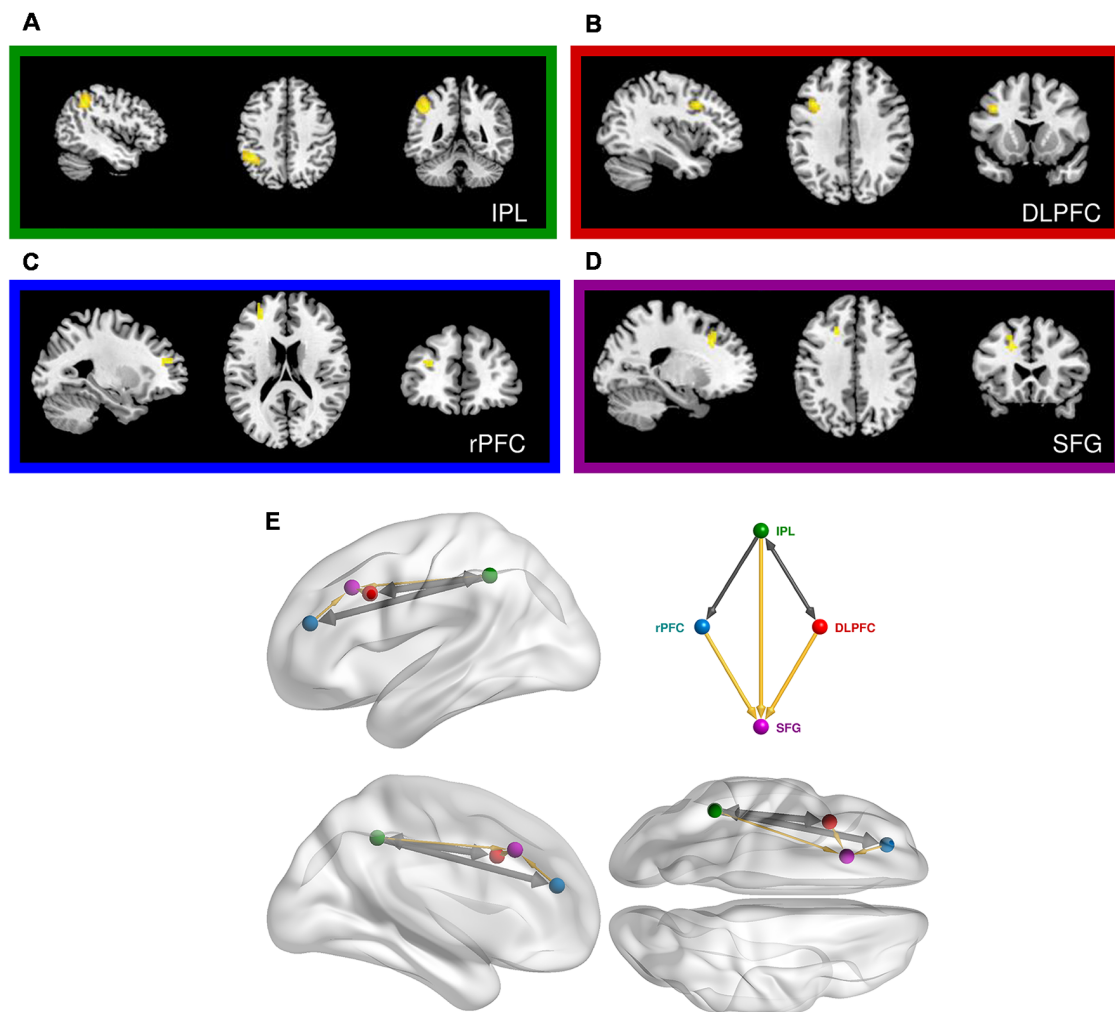
FIGURE 4 | Prefrontal cortical regions engaged in the same-different vibrotactile trials. **(A)** Main effect of the "Different > Same" contrast in the left inferior parietal lobe (IPL). **(B)** Main effect of the "Different > Same" contrast in the left dorsolateral prefrontal cortex (DLPFC). **(C)** Main effect of "Noise-free > Noisy" contrast in the left rostral PFC (rPFC). **(D)** Interaction of noise and difference in the left superior frontal gyrus (SFG). **(E)** Relative anatomical location of the corresponding nodes employed in the dynamic causal modeling (DCM), as labeled and colored in the inset. Thick arrows show effective connectivity common to all DCM models. Thin yellow links show connections used in some but not all models.

en route to the SFG. In parallel models, there is a direct effective connection from the IPL to the SFG in parallel to the rPFC and DLPFC connections. Additional modulatory influences are introduced on top of these architectures in order to explain the interaction effect in the SFG. In the nonlinear models (**Figure 5**, lower row) the modulation of inputs to SFG is mediated by modulation of connections from one area by another (namely DLPFC or rPFC). This activity-dependent modulation can be considered hierarchical. In contrast, in bilinear models (**Figure 5**, top row), this modulation is attributed directly to experimental inputs (namely, stimulus difference and regularity). In short, both bilinear and nonlinear models allow for context or state-dependent changes in afferents to the SFG: however, nonlinear models consider this state-dependent modulation to be dynamic and activity-dependent. These seven models encompass all possible such serial, parallel and hierarchical arrangements

considered separately. Because we sought a parsimonious and non-redundant model space, we did not consider models that combine these basic features (for example both serial and parallel connections).

BMS identified the double nonlinear and hierarchical model "Stork 3" as the model with the highest posterior exceedance probability of the seven tested (**Figure 6**). This model was followed by the other nonlinear models "Stork 2," and "Stork 1." The remaining bilinear models embodying serial and parallel motifs performed poorly as they were associated with a considerably lower exceedance probability (**Figure 6A**).

DISCUSSION

While being formed, stimulus representations contend with noise in the nervous system, placing an upper bound on
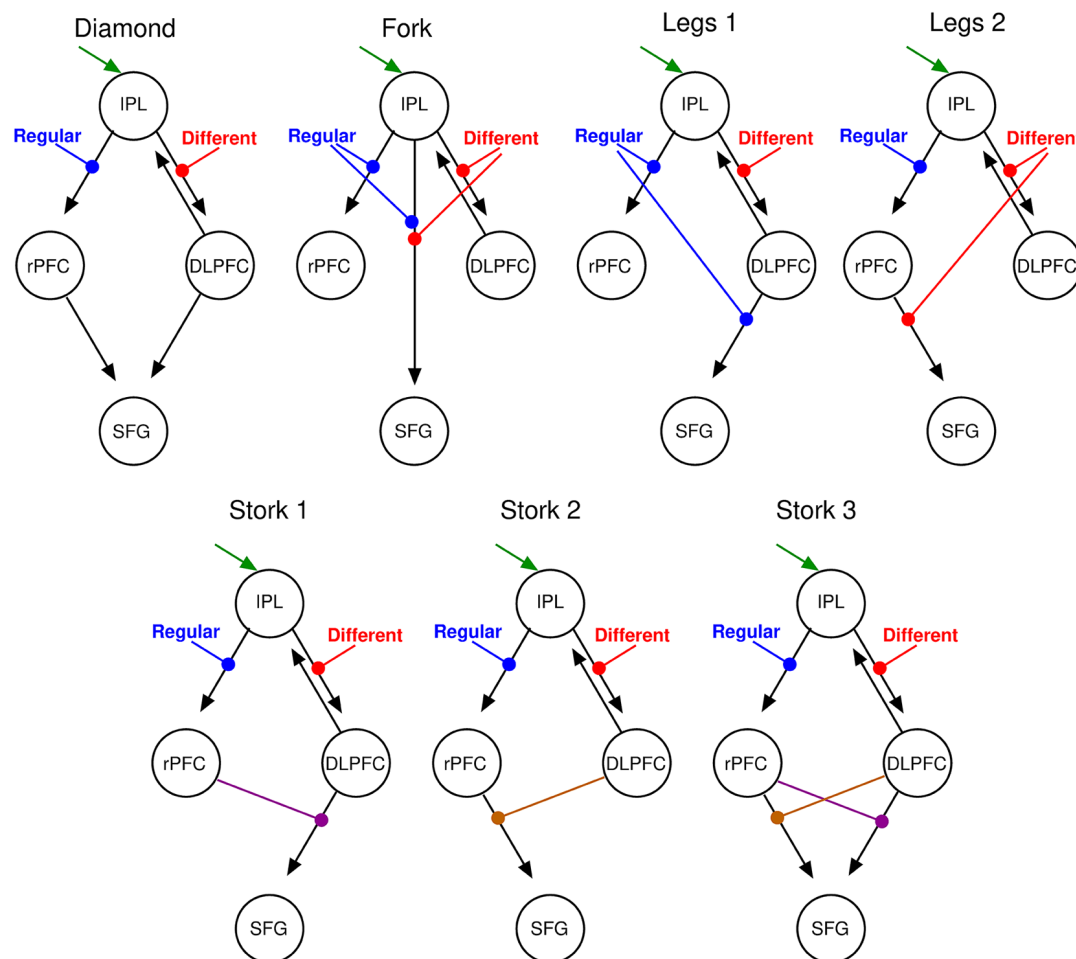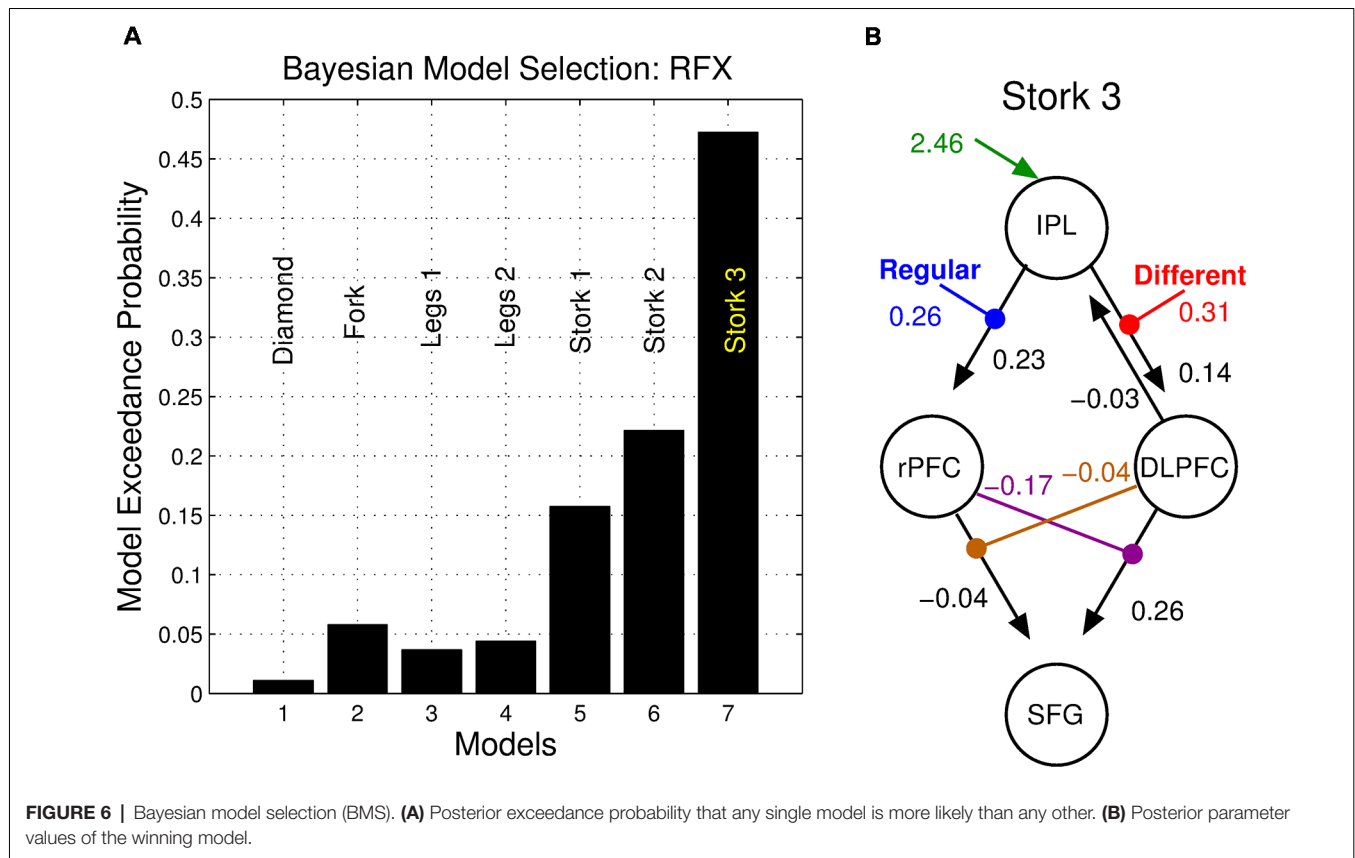
FIGURE 5 | DCM parsimonious and non-redundant model space. Stimulus inputs arrive *via* the IPL (green arrow) and propagate, *via* intrinsic connections (black arrows) to the rPFC and the dorsolateral PFC (DLPFC). Each of these intrinsic connections is perturbed by experimental inputs: different (red) and Regular (blue) stimulus trials that account for the corresponding effects in the SPM contrasts. From left to right, top motifs are linear serial (Diamond), parallel (Fork), hierarchical with the regular modulation at the higher level (Legs 1), and a hierarchical with the different modulation at the higher level (Legs 2). Bottom motifs are hierarchical and nonlinear. From left to right, regular modulation is at the top of the hierarchy (Stork 1), different modulation is at the top of the hierarchy (Stork 2), and the double non-linear model in which both modulations occupy top and low hierarchy positions at the different interactions (Stork 3). Please refer to the **Supplementary Material** for further discussion on the model space.

the precision of the stimulus representation and confounding any imprecision arising from the properties of the stimulus (Faisal et al., 2008). The precision of the ensuing percept is thus a composite of the stimulus noise and stochastic process in the perceptual system. This is crucial to perceptual inference: not only do we integrate information across modalities by weighting according to relative precision (Jacobs, 1999; Ernst et al., 2000), precision also plays a crucial role in combining new sensory evidence with prior knowledge to inform perceptual beliefs (Friston et al., 1996). However, there must also be a lower bound on precision in many everyday tasks, such that objects that are re-encountered can be recognized as familiar and, conversely, salience can be directed toward novel or surprising parts of the sensorium (Vossel et al., 2014). The modulation of factors influencing perceptual precision is thus context-dependent and under executive

control. Using a vibrotactile discrimination task whereby participants made contextual judgments that either implicitly required encoding of a precision estimate (same-different) or not (faster-slower), we identified a constellation of cortical regions predominantly in the left PFC that are engaged in computing, representing and deploying perceptual precision in the service of decision making. By modeling these effects, we observe that effective connectivity amongst these regions is subserved by a hierarchical network whereby activity in left rPFC and DLPFC exert a mutual gating influence on the SFG.

Accuracy is higher and responses are faster for simple magnitude comparisons (fast-slow) than during the detection of difference (same-different). As described by signal detection theory (MacMillan and Creelman, 2005), these two tasks differ in the way stimuli and noise are perceptually represented

FIGURE 6 | Bayesian model selection (BMS). **(A)** Posterior exceedance probability that any single model is more likely than any other. **(B)** Posterior parameter values of the winning model.

in "decision space": although perceptual uncertainty clearly plays a role in all decisions in our experiment (both faster-slower and same-different), the former task can be achieved simply by subtracting the inferred stimulus frequencies. By contrast, in the latter task, perceptual precision is explicitly part of the decision process, so that the perceived magnitude difference is weighed against the precision of each representation (**Figures 2**, **3**). This additional computational burden is reflected in slower reaction times (**Figure 3A**); the corresponding contextual functional neuroimaging contrast yielded a robust effect in the left IFG pars triangularis (BA 45), which lies within the mid VLPFC and has been implicated in the cognitive control of working memory (Badre and Wagner, 2007), a necessary component of our task. It has also been argued that the mid-VLPFC is involved in the "active retrieval" of information from posterior cortical association areas: active retrieval is required when stimuli in memory "do not bear stable relations to each other and therefore retrieval cannot be automatically driven by strong, stable, and unambiguous stimulus or context relations" (Petrides, 2002). This argument recapitulates the notion that additional neuronal resources are called upon when the ambiguity of perceptual representation becomes an integral aspect of the task at hand and not a mere nuisance factor.

To further understand the neural correlates of perceptual precision, we studied the consequence of degrading the temporal structure of the stimuli, thereby introducing controlled stimulus

noise. The contrast of regular > noisy trials in the same-different context showed additional activity in the left rPFC (BA 10, **Figure 4**), an apex region of the PFC. The rPFC has been associated with a broad variety of executive and integrative functions, including those that pertain to decision making (Koechlin and Hyafil, 2007; Li and Yang, 2012), working memory (Ramnani and Owen, 2004) and context (Simons et al., 2005). The stronger engagement of this region during the regular trials may be indicative of a requirement to account for the relatively high precision of stimulus representations arising from regular vibrations. This might reflect a fundamental role for this region in modifying perceptual stability to optimize the detection of change and surprise (Friston et al., 2012). Greater activity in regular compared to noisy vibrotactile stimuli has been previously observed in other regions of PFC during the explicit detection of stimulus noise (Godde et al., 2010). In our study, detecting the presence of noise was not explicitly required (or reported) but rather an implicit component of task execution. The rPFC may, therefore, encode a generic means of representing perceptual precision rather than a role linked specifically to explicit stimulus decoding. We return to this issue below.

The presence of noisy stimuli in the same-different task was either a help or a hindrance to task performance, depending upon the nature of the trial: consistent with our framing of decision-making in the presence of noise (**Figure 2**), noise increased the accuracy for same but not different trials.

In the case of same trials, stimulus noise may diminish the significance of the slight perception of difference that inevitably arises when encoding stimuli, even when such stimuli are physically identical. The presence of noise thus decreases the chance that such trials are mistakenly classified as different. However, the lower precision also increases the likelihood that the perception of difference associated with truly different trials is rendered subthreshold, increasing their misclassification. This behavioral interaction thus speaks directly to perceptual precision. The corresponding interaction contrast in our functional magnetic resonance imaging (fMRI) data yielded a cluster deep in the sulcus of left DLPFC cortex—the SFG. This finding suggests that in concert with other prefrontal regions such as the rPFC, the SFG may accumulate multiple aspects of decision-relevant evidence and integrate these on the fly.

We employed DCM to model dynamic network computations enacting the interaction of stimulus change and perceptual noise. The key features of the winning model (Stork 3) are nonlinear and hierarchical relationships between the DLPFC, the rPFC and the SFG (**Figure 5**). The balanced nature of this motif's structure mirrors the notion that the assessments of precision and stimulus difference mandate a mutual, dynamic exchange during the corresponding same-different task: high values of perceptual precision up-regulate the appreciation of stimulus change and likewise, the perception of change influences the role of precision on decisions. The nonlinear terms that account for the interaction effect ostensibly have an underlying biological basis—a "gating" mechanism, whereby the effective influence of activity from one neural region to another depends on the current activity in a third region. Candidate neural processes capable of underlying this effect include priming of voltage-dependent N-Methyl-D-aspartate (NMDA) channels through partial depolarization by AMPA-mediated synapses, synaptic depression/facilitation or early long-term potentiation (for review, see Stephan et al., 2008). The neural response of the SFG may thus depend on the immediate history of responses of the rPFC (facilitated by regular stimuli) and the DLPFC (facilitated stimulus difference), each influencing the other's concurrent influence.

The hierarchical organization of networks and information flow has been frequently described across prefrontal regions (Nee and D'Esposito, 2016). The "action-perception cycle" describes the complementary interaction between prefrontal networks of executive memory with a posterior network of perceptual memory, exerting reciprocal influences. This interaction is thought to occur at all levels of the nervous system, engaging neural networks at every hierarchical level of the neocortex (Fuster, 2009). All stages of processing generate internal feedback upon earlier stages, serving to monitor and modulate incoming signals at every stage (Fuster, 2006). Here, we have focused only on the interactions among the constellation of PFC regions identified by the task contrasts. The PFC is thought to constitute the highest level of the cortical hierarchy dedicated to the representation and execution of actions (Fuster, 2001). The analysis of functional and

structural hierarchies in PFC is a very active area of research (see Gorbach et al., 2011): to the best of our knowledge, this is the first study of hierarchies of effective connectivity within the human PFC underlying perceptual precision. The predominance of left PFC in this study may be partly due to the fact that all participants in our study were right handed and all stimuli were presented to the right index finger. The lateralization may thus be a consequence of the right-sided stimulus presentation rather than a reflection of hemispheric specialization. Most of our effects were indeed bilateral, although often only exceeding threshold in the left hemisphere (results not shown). Future work could also incorporate premotor regions involved in the task, likely in pre-empting the motor response.

It is important to note that the fast-slow < same-different contrast did not contain the same trials required for the same-different task. Hence, the full stimulus-set used by participants to set their decision-criteria in the same-different context is not present in this contrast. In addition to a substantially lower sensitivity (**Supplementary Figure S1**), participants possibly adopted a response bias towards responding "same" for the same-different context, reflected in higher accuracy (using proportion correct) for same trials than for different trials. Therefore, the fast-slow < same-different contrast examined in this study, whilst avoiding any confounds due to stimulus differences, is an incomplete comparison of stimulus representation between the two judgments. The neural regions identified from the contrast (IFG pars triangularis and middle temporal gyrus, **Supplementary Figure S1**) necessarily reflect the perceptual representation of the same-different judgment, and the computational criteria that underlies response bias.

We have framed the performance of our perceptual decision-making task in terms of Bayesian inference, i.e., that decisions depend upon weighting sensory evidence according to perceptual precision (Dayan et al., 1995; Karim et al., 2012). While all percepts accordingly involve both the perceptual value (mean) and the precision, our findings elucidate the manner in which this evidence and its precision are represented and integrated in a hierarchical prefrontal network when required for decision-making. For example, the representation of perceptual precision is associated with greater activity in the rPFC which then gates the effect of other stimulus properties. Our findings build on prior work regarding gain-mediated precision-weighted perceptual inference (Moran et al., 2013) and are consistent with the notion that neuronal activity encodes probability distributions regarding sensory evidence (Dayan et al., 1995; Sanger, 1996; Zemel et al., 1998). However, the application of classic DCM to fMRI data is limited to inferences regarding changes in local mean firing rates. Probabilistic population encoding likely also involves other moments of population activity, such as a direct mapping between the variance of neuronal states and the uncertainty of the perceptual representation (Beck et al., 2008; Shi and Griffiths, 2009). Although there exists a theoretical link between the variance of local population activity and gain control (Marreiros et al.,

2010), future work that employs stochastic variants of DCM (Li et al., 2011) could be used to infer higher order moments of neuronal activity (Harrison et al., 2005; Breakspear, 2013) and thus more directly probe the local neural correlates of perceptual precision.

ETHICS STATEMENT

Participants gave written informed consent and the study was approved by the University of New South Wales Human Research Ethics Committee.

AUTHOR CONTRIBUTIONS

LG, MK, JH, JM, and MB designed the research and wrote the manuscript. LG. MK, JH, and MB analyzed the data. LG, MK, and MB prepared the figures.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fncir. 2019.00027/full#supplementary-material

REFERENCES

Bach, D. R., and Dolan, R. J. (2012). Knowing how much you don't know: a neural organization of uncertainty estimates. *Nat. Rev. Neurosci.* 13, 572–586. doi: 10.1038/nrn3289

Badre, D., and Wagner, A. D. (2007). Left ventrolateral prefrontal cortex and the cognitive control of memory. *Neuropsychologia* 45, 2883–2901. doi: 10.1016/j.neuropsychologia.2007.06.015

Barak, O., Tsodyks, M., and Romo, R. (2010). Neuronal population coding of parametric working memory. *J. Neurosci.* 30, 9424–9430. doi: 10.1523/JNEUROSCI.1875-10.2010

Beck, J. M., Ma, W. J., Kiani, R., Hanks, T., Churchland, A. K., Roitman, J., et al. (2008). Probabilistic population codes for Bayesian decision making. *Neuron* 60, 1142–1152. doi: 10.1016/j.neuron.2008.09.021

Behrens, T. E., Woolrich, M. W., Walton, M. E., and Rushworth, M. F. (2007). Learning the value of information in an uncertain world. *Nat. Neurosci.* 10, 1214–1221. doi: 10.1038/nn1954

Botvinick, M. M. (2008). Hierarchical models of behavior and prefrontal function. *Trends Cogn. Sci.* 12, 201–208. doi: 10.1016/j.tics.2008.02.009

Breakspear, M. (2013). Dynamic and stochastic models of neuroimaging data: a comment on Lohmann et al. *Neuroimage* 75, 270–274. doi: 10.1016/j.neuroimage.2012.02.047

Britten, K. H., Shadlen, M. N., Newsome, W. T., and Movshon, J. A. (1992). The analysis of visual motion: a comparison of neuronal and psychophysical performance. *J. Neurosci.* 12, 4745–4765. doi: 10.1523/JNEUROSCI.12-12-04745.1992

Daw, N. D., Niv, Y., and Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat. Neurosci.* 8, 1704–1711. doi: 10.1038/nn1560

Dayan, P., Hinton, G. E., Neal, R. M., and Zemel, R. S. (1995). The helmholtz machine. *Neural Comput.* 7, 889–904. doi: 10.1162/neco.1995.7.5.889

Eickhoff, S. B., Stephan, K. E., Mohlberg, H., Grefkes, C., Fink, G. R., Amunts, K., et al. (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *Neuroimage* 25, 1325–1335. doi: 10.1016/j.neuroimage.2004.12.034

Engel, T. A., and Wang, X. J. (2011). Same or different? A neural circuit mechanism of similarity-based pattern match decision making. *J. Neurosci.* 31, 6982–6996. doi: 10.1523/JNEUROSCI.6150-10.2011

Ernst, M. O., and Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* 415, 429–433. doi: 10.1038/415429a

Ernst, M. O., Banks, M. S., and Bülthoff, H. H. (2000). Touch can change visual slant perception. *Nat. Neurosci.* 3, 69–73. doi: 10.1038/71140

Fairhall, S. L., and Ishai, A. (2007). Effective connectivity within the distributed cortical network for face perception. *Cereb. Cortex* 17, 2400–2406. doi: 10.1093/cercor/bhl148

Faisal, A. A., Selen, L. P., and Wolpert, D. M. (2008). Noise in the nervous system. *Nat. Rev. Neurosci.* 9, 292–303. doi: 10.1038/nrn2258

Fiorillo, C. D., Tobler, P. N., and Schultz, W. (2003). Discrete coding of reward probability and uncertainty by dopamine neurons. *Science* 299, 1898–1902. doi: 10.1126/science.1077349

Friston, K. (2005). A theory of cortical responses. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360, 815–836. doi: 10.1098/rstb.2005.1622

Friston, K. J., Breakspear, M., and Deco, G. (2012). Perception and self-organized instability. *Front. Comput. Neurosci.* 6:44. doi: 10.3389/fncom.2012.00044

Friston, K. J., Harrison, L., and Penny, W. (2003). Dynamic causal modelling. *Neuroimage* 19, 1273–1302. doi: 10.1016/S1053-8119(03)00202-7

Friston, K. J., Holmes, A., Poline, J.-B., Price, C. J., and Frith, C. D. (1996). Detecting activations in PET and fMRI: levels of inference and power. *Neuroimage* 4, 223–235. doi: 10.1006/nimg.1996.0074

Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J. P., Frith, C. D., and Frackowiak, R. S. (1994a). Statistical parametric maps in functional imaging: a general linear approach. *Hum. Brain Mapp.* 2, 189–210. doi: 10.1002/hbm.460020402

Friston, K. J., Worsley, K. J., Frackowiak, R. S., Mazziotta, J. C., and Evans, A. C. (1994b). Assessing the significance of focal activations using their spatial extent. *Hum. Brain Mapp.* 1, 210–220. doi: 10.1002/hbm.460010306

Fuster, J. M. (2001). The prefrontal cortex—an update: time is of the essence. *Neuron* 30, 319–333. doi: 10.1016/s0896-6273(01)00285-9

Fuster, J. M. (2006). The cognit: a network model of cortical representation. *Int. J. Psychophysiol.* 60, 125–132. doi: 10.1016/j.ijpsycho.2005.12.015

Fuster, J. M. (2009). Cortex and memory: emergence of a new paradigm. *J. Cogn. Neurosci.* 21, 2047–2072. doi: 10.1162/jocn.2009.21280

Godde, B., Diamond, M. E., and Braun, C. (2010). Feeling for space or for time: task-dependent modulation of the cortical representation of identical vibrotactile stimuli. *Neurosci. Lett.* 480, 143–147. doi: 10.1016/j.neulet.2010.06.027

Gold, J. I., and Shadlen, M. N. (2007). The neural basis of decision making. *Annu. Rev. Neurosci.* 30, 535–574. doi: 10.1146/annurev.neuro.29.051605.113038

Gorbach, N. S., Schütte, C., Melzer, C., Goldau, M., Sujazow, O., Jitsev, J., et al. (2011). Hierarchical information-based clustering for connectivity-based cortex parcellation. *Front. Neuroinform.* 5:18. doi: 10.3389/fninf.2011.00018

Grefkes, C., Nowak, D. A., Eickhoff, S. B., Dafotakis, M., Küst, J., Karbe, H., et al. (2008). Cortical connectivity after subcortical stroke assessed with functional magnetic resonance imaging. *Ann. Neurol.* 63, 236–246. doi: 10.1002/ana.21228

Harris, J. A. (2006). "Psychophysical investigations into cortical encoding of vibrotactile stimuli," in *Novartis Foundation Symposium*, (New York,

NY: John Wiley), 238. https://books.google.com.au/books?hl=en&lr=&id=EcA1avPlxZUC&oi=fnd&pg=PA238&dq=harris+2006+psychophysical+investigations+into+cortical+encoding+&ots=MM_DsMb-2l&sig=ojec0ihhJJaRFGEYEhXH-oNeSMU#v=onepage&q=harris%202006%20psychophysical%20investigations%20into%20cortical%20encoding&f=false

Harris, J. A., Arabzadeh, E., Fairhall, A. L., Benito, C., and Diamond, M. E. (2006). Factors affecting frequency discrimination of vibrotactile stimuli: implications for cortical encoding. *PLoS One* 1:e100. doi: 10.1371/journal.pone. 0000100

Harris, J. A., Miniussi, C., Harris, I. M., and Diamond, M. E. (2002). Transient storage of a tactile memory trace in primary somatosensory cortex. *J. Neurosci.* 22, 8720–8725. doi: 10.1523/JNEUROSCI.22-19-08720.2002

Harrison, L., David, O., and Friston, K. (2005). Stochastic models of neuronal dynamics. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360, 1075–1091. doi: 10.1098/rstb.2005.1648

Heekeren, H. R., Marrett, S., and Ungerleider, L. G. (2008). The neural systems that mediate human perceptual decision making. *Nat. Rev. Neurosci.* 9, 467–479. doi: 10.1038/nrn2374

Hegner, Y. L., Saur, R., Veit, R., Butts, R., Leiberg, S., Grodd, W., et al. (2007). BOLD adaptation in vibrotactile stimulation: neuronal networks involved in frequency discrimination. *J. Neurophysiol.* 97, 264–271. doi: 10.1152/jn.00 617.2006

Hernández, A., Zainos, A., and Romo, R. (2000). Neuronal correlates of sensory discrimination in the somatosensory cortex. *Proc. Natl. Acad. Sci. U S A* 97, 6191–6196. doi: 10.1073/pnas.120018597

Hsu, M., Bhatt, M., Adolphs, R., Tranel, D., and Camerer, C. F. (2005). Neural systems responding to degrees of uncertainty in human decision-making. *Science* 310, 1680–1683. doi: 10.1126/science.1115327

Huettel, S. A., Stowe, C. J., Gordon, E. M., Warner, B. T., and Platt, M. L. (2006). Neural signatures of economic preferences for risk and ambiguity. *Neuron* 49, 765–775. doi: 10.1016/j.neuron.2006.01.024

Jacobs, R. A. (1999). Optimal integration of texture and motion cues to depth. *Vision Res.* 39, 3621–3629. doi: 10.1016/s0042-6989(99)00088-7

Kalberlah, C., Villringer, A., and Pleger, B. (2013). Dynamic causal modeling suggests serial processing of tactile vibratory stimuli in the human somatosensory cortex—an fMRI study. *Neuroimage* 74, 164–171. doi: 10.1016/j.neuroimage.2013.02.018

Karim, M., Harris, J. A., Morley, J. W., and Breakspear, M. (2012). Prior and present evidence: how prior experience interacts with present information in a perceptual decision making task. *PLoS One* 7:e37580. doi: 10.1371/journal. pone.0037580

Kayser, A. S., Buchsbaum, B. R., Erickson, D. T., and D'Esposito, M. (2010). The functional anatomy of a perceptual decision in the human brain. *J. Neurophysiol.* 103, 1179–1194. doi: 10.1152/jn.00364.2009

Kennerley, S. W., Walton, M. E., Behrens, T. E., Buckley, M. J., and Rushworth, M. F. (2006). Optimal decision making and the anterior cingulate cortex. *Nat. Neurosci.* 9, 940–947. doi: 10.1038/nn1724

Knill, D. C., and Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends Neurosci.* 27, 712–719. doi: 10.1016/j. tins.2004.10.007

Koechlin, E., and Hyafil, A. (2007). Anterior prefrontal function and the limits of human decision-making. *Science* 318, 594–598. doi: 10.1126/science.11 42995

Leff, A. P., Schofield, T. M., Stephan, K. E., Crinion, J. T., Friston, K. J., and Price, C. J. (2008). The cortical dynamics of intelligible speech. *J. Neurosci.* 28, 13209–13215. doi: 10.1523/JNEUROSCI.2903-08.2008

Li, B., Daunizeau, J., Stephan, K. E., Penny, W., Hu, D., and Friston, K. (2011). Generalised filtering and stochastic DCM for fMRI. *Neuroimage* 58, 442–457. doi: 10.1016/j.neuroimage.2011.01.085

Li, S., and Yang, F. (2012). Task-dependent uncertainty modulation of perceptual decisions in the human brain. *Eur. J. Neurosci.* 36, 3732–3739. doi: 10.1111/ejn. 12006

MacMillan, N. A., and Creelman, C. D. (2005). *Detection Theory: A User's Guide.* 2nd Edn. Mahwah, NJ: Psychology Press.

Marreiros, A. C., Stephan, K. E., and Friston, K. J. (2010). Dynamic causal modeling. *Scholarpedia* 5:9568. doi: 10.4249/scholarpedia.9568

Mechelli, A., Price, C. J., Noppeney, U., and Friston, K. J. (2003). A dynamic causal modeling study on category effects: bottom-up or top-down mediation? *J. Cogn. Neurosci.* 15, 925–934. doi: 10.1162/089892903770007317

Mesulam, M.-M. (1998). From sensation to cognition. *Brain* 121, 1013–1052. doi: 10.1093/brain/121.6.1013

Miller, P., Brody, C. D., Romo, R., and Wang, X. J. (2003). A recurrent network model of somatosensory parametric working memory in the prefrontal cortex. *Cereb. Cortex* 13, 1208–1218. doi: 10.1093/cercor/bhg101

Miller, E. K., and Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.* 24, 167–202. doi: 10.1146/annurev.neuro. 24.1.167

Moran, R. J., Campo, P., Symmonds, M., Stephan, K. E., Dolan, R. J., and Friston, K. J. (2013). Free energy, precision and learning: the role of cholinergic neuromodulation. *J. Neurosci.* 33, 8227–8236. doi: 10.1523/jneurosci.4255-12. 2013

Navajas, J., Hindocha, C., Foda, H., Keramati, M., Latham, P. E., and Bahrami, B. (2017). The idiosyncratic nature of confidence. *Nat. Hum. Behav.* 1, 810–818. doi: 10.1038/s41562-017-0215-1

Nee, D. E., and D'Esposito, M. (2016). The hierarchical organization of the lateral prefrontal cortex. *eLife* 5:e12112. doi: 10.7554/elife.12112

Noppeney, U., Josephs, O., Hocking, J., Price, C. J., and Friston, K. J. (2008). The effect of prior visual information on recognition of speech and sounds. *Cereb. Cortex* 18, 598–609. doi: 10.1093/cercor/bhm091

Petersen, S. E., and Sporns, O. (2015). Brain networks and cognitive architectures. *Neuron* 88, 207–219. doi: 10.1016/j.neuron.2015.09.027

Petrides, M. (2002). The mid-ventrolateral prefrontal cortex and active mnemonic retrieval. *Neurobiol. Learn. Mem.* 78, 528–538. doi: 10.1006/nlme.2002.4107

Pleger, B., Ruff, C. C., Blankenburg, F., Bestmann, S., Wiech, K., Stephan, K. E., et al. (2006). Neural coding of tactile decisions in the human prefrontal cortex. *J. Neurosci.* 26, 12596–12601. doi: 10.1523/jneurosci.4275-06.2006

Pouget, A., Beck, J. M., Ma, W. J., and Latham, P. E. (2013). Probabilistic brains: knowns and unknowns. *Nat. Neurosci.* 16, 1170–1178. doi: 10.1038/nn.3495

Preuschhof, C., Heekeren, H. R., Taskin, B., Schubert, T., and Villringer, A. (2006). Neural correlates of vibrotactile working memory in the human brain. *J. Neurosci.* 26, 13231–13239. doi: 10.1523/jneurosci.2767-06.2006

Ramnani, N., and Owen, A. M. (2004). Anterior prefrontal cortex: insights into function from anatomy and neuroimaging. *Nat. Rev. Neurosci.* 5, 184–194. doi: 10.1038/nrn1343

Romo, R., and Salinas, E. (2003). Flutter discrimination: neural codes, perception, memory and decision making. *Nat. Rev. Neurosci.* 4, 203–218. doi: 10.1038/nrn1058

Rushworth, M. F., and Behrens, T. E. (2008). Choice, uncertainty and value in prefrontal and cingulate cortex. *Nat. Neurosci.* 11, 389–397. doi: 10.1038/nn2066

Sanger, T. D. (1996). Probability density estimation for the interpretation of neural population codes. *J. Neurophysiol.* 76, 2790–2793. doi: 10.1152/jn.1996.76. 4.2790

Shi, L., and Griffiths, T. L. (2009). "Neural implementation of hierarchical Bayesian inference by importance sampling," in *Advances in Neural Information Processing Systems*, eds Y. Bengio, D. Schuurmans, J. Lafferty, C. Williams and A. Culotta (Cambridge, MA: MIT Press), 1669–1677.

Simons, J. S., Owen, A. M., Fletcher, P. C., and Burgess, P. W. (2005). Anterior prefrontal cortex and the recollection of contextual information. *Neuropsychologia* 43, 1774–1783. doi: 10.1016/j.neuropsychologia.2005.02.004

Smith, A. P., Stephan, K. E., Rugg, M. D., and Dolan, R. J. (2006). Task and content modulate amygdala-hippocampal connectivity in emotional retrieval. *Neuron* 49, 631–638. doi: 10.1016/j.neuron.2005.12.025

Sörös, P., Marmurek, J., Tam, F., Baker, N., Staines, W. R., and Graham, S. J. (2007). Functional MRI of working memory and selective attention in vibrotactile frequency discrimination. *BMC Neurosci.* 8:48. doi: 10.1186/1471-2202-8-48

Stephan, K. E., Kasper, L., Harrison, L. M., Daunizeau, J., den Ouden, H. E., Breakspear, M., et al. (2008). Nonlinear dynamic causal models for fMRI. *Neuroimage* 42, 649–662. doi: 10.1016/j.neuroimage.2008.04.262

Stephan, K. E., Marshall, J. C., Penny, W. D., Friston, K. J., and Fink, G. R. (2007). Interhemispheric integration of visual processing during task-driven lateralization. *J. Neurosci.* 27, 3512–3522. doi: 10.1523/jneurosci.4766-06.2007

Stephan, K. E., Penny, W. D., Daunizeau, J., Moran, R. J., and Friston, K. J. (2009). Bayesian model selection for group studies. *Neuroimage* 46, 1004–1017. doi: 10.1016/j.neuroimage.2009.03.025

Stephan, K. E., Penny, W. D., Moran, R. J., den Ouden, H. E., Daunizeau, J., and Friston, K. J. (2010). Ten simple rules for dynamic causal modeling. *Neuroimage* 49, 3099–3109. doi: 10.1016/j.neuroimage.2009.11.015

Summerfield, C., Egner, T., Greene, M., Koechlin, E., Mangels, J., and Hirsch, J. (2006). Predictive codes for forthcoming perception in the frontal cortex. *Science* 314, 1311–1314. doi: 10.1126/science.1132028

Summerfield, C., and Koechlin, E. (2008). A neural representation of prior information during perceptual inference. *Neuron* 59, 336–347. doi: 10.1016/j.neuron.2008.05.021

Tobler, P. N., O'Doherty, J. P., Dolan, R. J., and Schultz, W. (2007). Reward value coding distinct from risk attitude-related uncertainty coding in human reward systems. *J. Neurophysiol.* 97, 1621–1632. doi: 10.1152/jn.00745.2006

Vossel, S., Mathys, C., Daunizeau, J., Bauer, M., Driver, J., Friston, K. J., et al. (2014). Spatial attention, precision, and bayesian inference: a study of saccadic response speed. *Cereb. Cortex* 24, 1436–1450. doi: 10.1093/cercor/bhs418

Wang, X.-J. (2008). Decision making in recurrent neuronal circuits. *Neuron* 60, 215–234. doi: 10.1016/j.neuron.2008.09.034

Wang, X.-J. (2012). Neural dynamics and circuit mechanisms of decision-making. *Curr. Opin. Neurobiol.* 22, 1039–1046. doi: 10.1016/j.conb.2012.08.006

Yu, A. J., and Dayan, P. (2005). Uncertainty, neuromodulation, and attention. *Neuron* 46, 681–692. doi: 10.1016/j.neuron.2005.04.026

Zemel, R. S., Dayan, P., and Pouget, A. (1998). Probabilistic interpretation of population codes. *Neural Comput.* 10, 403–430. doi: 10.1162/089976698300017818

Zwislocki, J. J., and Relkin, E. M. (2001). On a psychophysical transformed-rule up and down method converging on a 75% level of correct responses. *Proc. Natl. Acad. Sci. U S A* 98, 4811–4814. doi: 10.1073/pnas.081082598

Uncovering the Transcriptional Correlates of Hub Connectivity in Neural Networks

Aurina Arnatkevičiūtė[1]*, Ben D. Fulcher[1,2] and Alex Fornito[1]

[1] Monash Biomedical Imaging, School of Psychological Sciences, Turner Institute for Brain and Mental Health, Monash University, Clayton, VIC, Australia, [2] School of Physics, The University of Sydney, Sydney, NSW, Australia

Connections in nervous systems are disproportionately concentrated on a small subset of neural elements that act as network hubs. Hubs have been found across different species and scales ranging from *C. elegans* to mouse, rat, cat, macaque, and human, suggesting a role for genetic influences. The recent availability of brain-wide gene expression atlases provides new opportunities for mapping the transcriptional correlates of large-scale network-level phenotypes. Here we review studies that use these atlases to investigate gene expression patterns associated with hub connectivity in neural networks and present evidence that some of these patterns are conserved across species and scales.

Keywords: connectome, hub, rich-club, gene expression, network neuroscience, graph theory, genome

1. INTRODUCTION

The brain is a multiscale network, with neuronal elements exhibiting coordinated patterns of activity that unfold across several orders of magnitude in time and space (Buzsáki and Draguhn, 2004; Lichtman and Denk, 2011; Fornito et al., 2016). Graph theory provides a useful approach to represent network organization at each scale by focusing on the essential elements of the system: processing units and their interactions, represented, respectively, as nodes and edges in the graph (Bullmore and Sporns, 2009; Fornito et al., 2016). The advantage of using a graph theoretic approach to understand the organizational properties of the brain is that the same analysis tools can be applied regardless of the species or scale, ranging from electron micrograph data of neuron-and-synapse connectivity in the nematode worm *Caenorhabditis elegans* (White et al., 1986; Varshney et al., 2011), through tract-tracing data in the mouse (Oh et al., 2014; Gămănuţ et al., 2018) and macaque (Stephan et al., 2001; Markov et al., 2014), to brain-wide non-invasive structural and functional imaging in the human (Bassett and Bullmore, 2009; Bullmore and Sporns, 2009; Fornito et al., 2013).

A growing body of work has demonstrated that the connection topology of neural networks—that is, the specific arrangement of connections between system elements—shows a number of non-random properties that are conserved across different scales and in different species (Bullmore and Sporns, 2009; Sporns, 2011; Fornito et al., 2016; van den Heuvel et al., 2016a; Schröter et al., 2017). These include (i) a predominance of short-range, locally clustered connections supporting functional specialization coupled with sparse, long-range projections that may promote global integration and functional diversity, resulting in an economical small-world organization (Watts and Strogatz, 1998; Bassett and Bullmore, 2017; Betzel and Bassett, 2017); (ii) the presence of densely connected sub-networks, termed modules, organized hierarchically across several

resolution levels so that modules contain nested sub-modules and so on (Meunier et al., 2009; Bassett et al., 2010); (iii) a fat-tailed distribution of connectivity across nodes, such that some nodes possess a relatively large number of connections and act as network hubs (van den Heuvel and Sporns, 2011; Towlson et al., 2013; van den Heuvel et al., 2016a); and (iv) a dense inter-connectivity of hub nodes, leading to the formation of a "rich-club" (Zamora-López et al., 2010; van den Heuvel and Sporns, 2011; Harriger et al., 2012; Towlson et al., 2013).

The strong conservation of such topological properties across scales and species implies that particular connectivity patterns are being evolutionary favored either through common descent or convergent evolutionary paths. This raises questions concerning the degree to which genes influence brain network topology. Twin studies have shown that topological properties of human brain networks mapped at the macroscale are heritable (Smit et al., 2008; Fornito et al., 2011; van den Heuvel et al., 2013a; Bohlken et al., 2014; Sinclair et al., 2015; Zhan et al., 2015; Colclough et al., 2017), but they do not indicate the specific genes involved. Studies linking structural variation in the genome to variability in network-level phenotypes, both at the level of candidate genes (Liu et al., 2010; Brown et al., 2011; Dennis et al., 2011; Markett et al., 2017) and in genome-wide scans (Jahanshad et al., 2013), have started to address this gap. However, they provide a partial picture, as it is often unclear how a given DNA variant impacts gene expression to give rise to phenotypic variability.

In neuroscience, it has been difficult to link direct measures of gene expression to variation in network phenotypes defined across large swathes of the brain, as gene expression has traditionally only been quantifiable though invasive interrogation of regionally localized tissue samples. The recent availability of large-scale, brain-wide atlases of gene expression (Lein et al., 2007; Hawrylycz et al., 2012), has overcome this hurdle and presented new opportunities to understand the molecular correlates of network-level phenotypes. Patterns of gene expression have been used to predict whether two neurons (or large-scale brain regions) will be structurally connected (Kaufman et al., 2006; Varadan et al., 2006; Baruch et al., 2008; French and Pavlidis, 2011; Wolf et al., 2011; Ji et al., 2014; Fakhry and Ji, 2015), and confirmed that regional variations in gene expression track specific aspects of structural (Goel et al., 2014; Forest et al., 2017; Parkes et al., 2017; Romero-Garcia et al., 2018) and functional (Cioli et al., 2014; Hawrylycz et al., 2015; Richiardi et al., 2015; Krienen et al., 2016; Anderson et al., 2018) brain networks. The integration of gene expression atlases with imaging data is also shedding light on the molecular correlates of macroscopic brain changes observed in a range of disorders, such as Huntington's disease (McColgan et al., 2018), Parkinson's disease (Rittman et al., 2016), and schizophrenia (Romme et al., 2017).

One important aspect of brain network organization is the distribution of connections across nodes, which is disproportionately concentrated on a small number of network hubs (van den Heuvel and Sporns, 2011; Towlson et al., 2013). Most simply, network hubs are defined as nodes with a relatively large number of connections, placing them in a topologically

central position within the network (although other definitions are possible; see Power et al., 2011; Oldham et al., 2018). Intuitively, the global air transportation network offers insight into the role of hubs in mediating network traffic flow; certain airports, such as Dubai International, London Heathrow, and LAX are linked to the rest of the network by a much larger number of direct flights than other airports. They are thus positioned to mediate a large fraction of intercontinental travel. Similarly, connections are not distributed equally across neurons, neuronal populations or large brain areas, with specific network elements possessing the lion's share of connections (van den Heuvel and Sporns, 2011; Towlson et al., 2013; de Reus and van den Heuvel, 2014; van den Heuvel et al., 2016a). These brain hubs are thought to play a critical role in the functional integration of anatomically disparate systems (Harriger et al., 2012; van den Heuvel et al., 2012), and are disproportionately impacted by a diverse variety of brain diseases (Crossley et al., 2014; Fornito et al., 2015). Thus, understanding the molecular basis for hub connectivity may provide insights not only into integrated cerebral function, but also into the various disease processes that plague the brain.

In this article, we review how brain-wide gene expression atlases have been used to link two traditionally disparate scales of analysis in neuroscience: molecular function (microscale) and whole-brain network topology (macroscale), by identifying the transcriptional correlates of brain network hubs. We begin with a brief overview of the expression atlases that are currently available and then consider how hubs are defined in brain networks and what we know about their functional role. We then examine research indicating that brain network hubs possess a distinct and conserved transcriptional signature.

2. CHARACTERIZING GENE EXPRESSION ACROSS THE ENTIRE BRAIN

Gene expression is a process through which genetic information encoded in sequences of DNA is read and used to synthesize a particular gene product. The two key steps in this complex process are transcription, where an unwound segment of DNA is read to produce messenger (mRNA), and translation, which occurs when the resulting mRNA is used to synthesize the gene product, such as a protein. Gene expression is commonly inferred from mRNA levels, thus serving as an index of transcriptional activity—an indirect proxy for the protein abundance. Transcriptional activity can be measured using several different techniques that either assay bulk tissue samples [microarray (Schulze and Downward, 2001), RNA-seq (Mortazavi et al., 2008; Wang et al., 2009)], histological sections at a cellular resolution [in situ hybridization (ISH) (Schulze and Downward, 2001)], or single cells [single-cell RNA sequencing (scRNA-seq) (Tang et al., 2009)]. Different classes of brain cells show distinctive gene expression patterns (Darmanis et al., 2015; Poulin et al., 2016; Tasic et al., 2016; Mancarci et al., 2017), and scRNA-seq is thus regarded as the most promising technology for accurately resolving cell-type specificity (Yu and Lin, 2016). However, scRNA-seq is difficult to scale to brain-wide analyses,

and current brain-wide atlases of gene expression have relied on microarray or ISH. ISH has high spatial resolution, allowing gene expression to be measured in a tissue section with relatively high sensitivity and specificity, but requires a very large number of samples to quantify expression levels across thousands of genes (Unger et al., 2010). ISH has therefore only been used to construct atlases for species with high tissue availability, such as the mouse (Lein et al., 2007). Microarray, on the other hand, allows the quantification of expression levels of thousands of genes at once by measuring the hybridization of cRNA (Cy3-labeled RNA) in a tissue sample to particular spot (probe) on the microarray chip. The technique is limited to known gene sequences and is prone to background noise (Okoniewski and Miller, 2006; Royce et al., 2007), but provides a cost-effective way to measure gene transcription in high-throughput manner. It has been used to produce spatially comprehensive atlases of the human (Kang et al., 2011; Hawrylycz et al., 2012; Miller et al., 2014) and non-human primate brain [NIH Blueprint Non-Human Primate (NHP) Atlas (2009), in conjunction with ISH].

As summarized in Keil et al. (2018), there is a large number of gene expression atlases. Due to their high spatial coverage, the two most used brain-wide expression atlases are the Allen Mouse Brain Atlas (AMBA) (Lein et al., 2007) and the Allen Human Brain Atlas (AHBA) (Hawrylycz et al., 2012), both made freely available by the Allen Institute for Brain Science. The AMBA provides an extensive representation of the expression patterns of 19 419 genes across the whole mouse brain, using ISH to quantify brain-wide expression patterns with the cellular resolution at each tissue slice with slices acquired every $200\mu m$ (the latter resolution depends on the section). Spatially resolved gene expression data can be further parcellated using anatomical atlases of the mouse brain (Johnson et al., 2010; Furth et al., 2018) to acquire averaged expression values through a hierarchy of brain regions defined at different resolution scales. The AHBA comprises expression measures for 21, 245 genes (depending on available annotation data) taken from 3, 702 spatially distinct post-mortem tissue samples distributed throughout the brains of six human donors (Hawrylycz et al., 2012, 2015). Both atlases have been mapped to stereotaxic space, allowing researchers to link spatial variations in gene expression to the spatial variations of a given neural phenotype (i.e., any quantifiable, spatially varying property of the brain, as measured either at the level of brain regions or pairs of regions) (Fornito et al., 2019). Other gene expression databases include both spatial (Fertuzinhos et al., 2014) and spatio-temporal (Ayoub et al., 2011; Belgard et al., 2011; Colantuoni et al., 2011; Miller et al., 2014) atlases, along with the Allen Developing Mouse Brain Atlas (2008), however most of these lack the spatial coverage of the AMBA and AHBA with only a handful regions being assessed across multiple time points. Some gene expression atlases have also been published for the macaque, using ISH and microarray (Bakken et al., 2016), and *C. elegans* (Harris et al., 2010). The latter database has been curated from published reports and contains binary entries on around 5% of the $\sim 20,000$ genes in the full worm genome, such that the only information encoded is whether a given gene is expressed or not in a neuron.

Gene expression measures can be influenced by a number of technical and biological factors (Fraser et al., 2005; Berchtold

et al., 2008; Kumar et al., 2013; Trabzuni et al., 2013). For example, the AHBA consists of data from six donor brains, each varying in characteristics, such as age at death, cause of death, sex, and ethnicity. Therefore, any analysis pooling expression measures across brains should ensure that inter-subject variability has not directly influenced the results. The analysis of gene expression measures often involves important additional processing decisions that are not applied consistently and can impact final results. For example, useful steps in processing raw AHBA data prior to analysis include (i) verifying probe-to-gene annotations; (ii) filtering genes that are not expressed above the background; (iii) selecting a representative probe when more than one probe has been used to assay a single gene; (iv) assigning tissue samples to specific brain regions in the imaging dataset; and (v) normalizing expression measures to account for inter-individual differences and outlying values. Each step requires a number of decisions, and best-practice workflows have not been established yet (Arnatkevičiūtė et al., 2019). Finally, gene expression data often shows a strong spatial autocorrelation, such that gene expression is more tightly coupled between regions that are close to each other compared to those that are spatially distant. This trend has been demonstrated in the mouse (Fulcher and Fornito, 2016), human (Richiardi et al., 2015; Krienen et al., 2016; Vértes et al., 2016; Pantazatos and Li, 2017; Arnatkevičiūtė et al., 2019) and head of *C. elegans* (Arnatkevičiūtė et al., 2018). In order to demonstrate that a putative association between regional variations in gene expression and a given neural phenotype is evident beyond this distance-dependence, potential biases introduced by the dependence can be addressed using methods ranging from simple regression (Fulcher and Fornito, 2016), partial Mantel tests (French and Pavlidis, 2011; Ji et al., 2014; Fakhry et al., 2015) or spatially constrained randomization procedures (for example, see Vértes et al., 2016; Burt et al., 2017; Seidlitz et al., 2018; Arnatkevičiūtė et al., 2019).

Brain-wide gene expression measures can be related to a brain network-level phenotype either at the level of specific brain regions (Myers et al., 2007; Rittman et al., 2016; Vértes et al., 2016; Parkes et al., 2017) or using inter-regional transcriptional coupling (Richiardi et al., 2015; Fulcher and Fornito, 2016; Arnatkevičiūtė et al., 2018; Romero-Garcia et al., 2018). Analyses of regional gene expression focus on understanding how the expression of a given gene varies across regions, and whether this variation tracks spatial variations in some other phenotype (e.g., regional gray matter volume, or number of connections). In analyses of inter-regional transcriptional coupling or correlated gene expression (CGE), each region's transcriptional profile is mapped as a vector of expression values across all genes, and these vectors are correlated between different regions, thus resulting in a region × region CGE matrix indicating the similarity between brain regions in terms of their gene expression patterns. Gene-to-gene co-expression (Eising et al., 2016; Keo et al., 2017; Negi and Guda, 2017), on the other hand, is estimated at the levels of genes (rather than regions). Each gene's expression profile across regions is summarized as a vector, and these vectors are correlated between pairs of genes, resulting in a gene × gene coexpression matrix demonstrating whether regional expression patterns for gene pairs match. Note that the term gene

coexpression is sometimes used in reference to CGE. We use the current nomenclature to avoid confusion between the two.

Once a relationship between gene expression and a given neural phenotype has been established, functional groups of genes involved in driving the effect can be identified using gene set enrichment analyses (GSEA) (Subramanian et al., 2005; Irizarry et al., 2009). Since such analyses are often performed across many thousands of genes, GSEA offers a method for determining whether certain categories of genes—e.g., defined by gene ontology (GO) (Ashburner et al., 2000) or KEGG ontology (KO) (Kanehisa and Goto, 2000)—are over-represented in the set of genes showing the strongest associations. This approach allows for a functional interpretation of the results, at the expense of specificity at the level of single genes (i.e., inferences are made about functional groups of genes).

3. HUBS IN BRAIN NETWORKS

Complex behaviors require the coordination and integration of information both within and across different, functionally specialized brain regions. In primate brains, it has long been assumed that association areas, sitting atop the cortical hierarchy, and in interaction with subcortical regions, play an important role in these integrative processes (Felleman and Van Essen, 1991; Mesulam, 1998; Meyer and Damasio, 2009). Structural connectivity studies have confirmed that association areas, and regions of basal ganglia and thalamus, have high levels of connectivity, marking them as network hubs (van den Heuvel and Sporns, 2011). Artificially lesioning these nodes rapidly fragments the network, indicating that they play a vital role in network integration (Albert et al., 2000; van den Heuvel and Sporns, 2011). Moreover, both simulated node deletion and *in vivo* regional inactivation experiments demonstrate a direct relationship between a brain region's centrality and its functional impact on connected networks (Vetere et al., 2017).

Network hubs, the core elements in the network, can be defined using a range of different measures. These measures quantify distinct aspects of topological centrality, which can be defined as the capacity of a node to influence or be influenced by other nodes by virtue of its connection topology (Fornito et al., 2016). The simplest such measure is node degree, which is defined as the number of connections attached to a node. Other commonly used measures include closeness and betweenness centrality, which are both built on the premise that information in the network propagates through the most efficient route (the shortest path between regions), and thus, the centrality of any given node can be quantified by its average shortest path length (closeness), or the number of shortest paths between other nodes on which it lies (betweenness). These measures are often positively correlated across most networks, including the brain, and it is common to find a subset of nodes that score highly on most centrality measures, representing a topologically central network core (Oldham et al., 2018).

Another way to define hubs is in relation to the modular organization of the network. Nodes within a module are densely interconnected with eac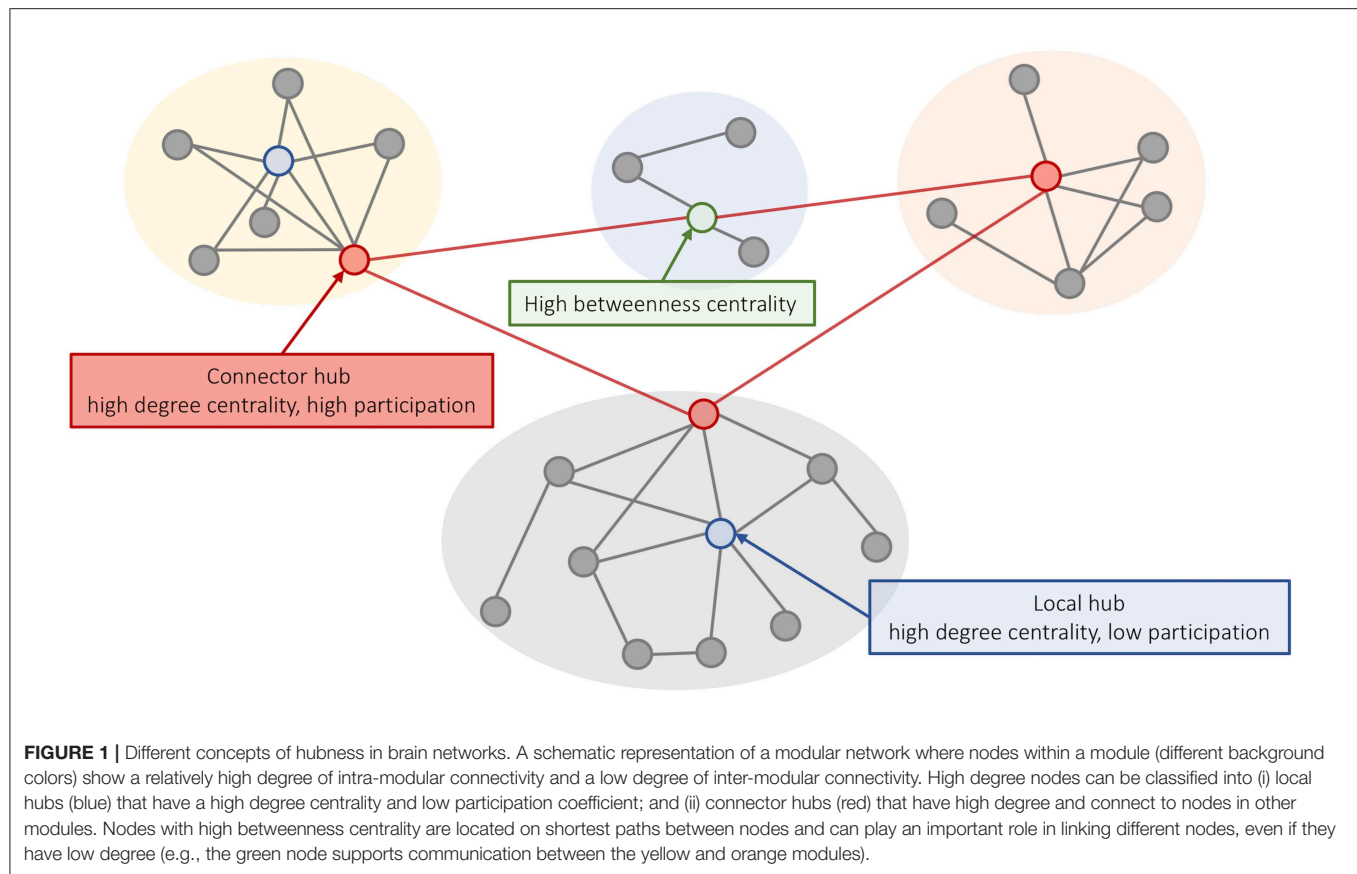h other and relatively sparsely connected to nodes in other modules. Given a partition of a network into modules (e.g., Blondel et al., 2008), the integrative role of a node in the network can be characterized using the participation coefficient: a measure of connection diversity that assigns a high score to nodes with connections distributed evenly across modules. Thus, hubs defined based on the degree centrality can be further classified into "local hubs," which connect primarily to nodes in the same module (high degree and low participation), and "connector hubs," which connect to nodes from other modules (**Figure 1**) (Guimerá et al., 2012).

The interpretation of different measures of network centrality must be moderated by an appreciation of how the network has been constructed. If one investigates structural connectivity (e.g., through electron microscopy, tract tracing, or diffusion MRI) then network edges represent physical connections between network elements, and interpretation is straightforward. If one investigates functional connectivity (e.g., through electrophysiology, calcium imaging, or functional MRI), which captures statistical dependencies between physiological signals recorded at each node (Friston, 1994), the interpretation is less clear and some measures of dependence, such as the correlation coefficient, can bias the topology of the network (Power et al., 2011; Zalesky et al., 2012). Furthermore, different centrality measures make assumptions about how dynamics unfold on the network structure. For example, closeness and betweenness assume information is routed along shortest paths, which may not be a realistic model of communication in nervous systems (Goñi et al., 2014; Mišić et al., 2015; Seguin et al., 2018).

Brain network hubs are densely interconnected, forming a rich-club (Colizza et al., 2006). This property has been observed in the macroscale human connectome (van den Heuvel and Sporns, 2011), the mesoscale connectomes of the mouse (Fulcher and Fornito, 2016), rat (van den Heuvel et al., 2016b), cat (de Reus and van den Heuvel, 2013) and macaque (Harriger et al., 2012), and the micro-scale neuronal connectome of the *C. elegans* (Towlson et al., 2013) (**Figure 2**).

Given that hubs are distributed throughout the brain and involved in diverse functional systems (de Reus and van den Heuvel, 2013; van den Heuvel and Sporns, 2013; Fulcher and Fornito, 2016), dense inter-connectivity of hub nodes is thought to support efficient integration of different functionally specialized systems (van den Heuvel et al., 2012), and to increase the diversity of the brain's functional repertoire (Senden et al., 2014). This integrative capacity comes at cost, with connections between hubs extending over longer anatomical distances than other types of connections (van den Heuvel and Sporns, 2011; Harriger et al., 2012; Fulcher and Fornito, 2016; Arnatkevičiūtė et al., 2018). Hub regions also have the highest levels of resting metabolism (Vaishnavi et al., 2010; Tomasi et al., 2013) and blood flow (Liang et al., 2013). This high metabolic cost is thought to partly explain why pathology preferentially accumulates in brain network hubs across a wide range of diverse neurological diseases (Bullmore and Sporns, 2012; Crossley et al., 2014; Fornito et al., 2015).

The mechanisms resulting in the emergence of network hubs are unknown, but geometric constraints and evolutionary pressures to maximize adaptive function may play a role

FIGURE 1 | Different concepts of hubness in brain networks. A schematic representation of a modular network where nodes within a module (different background colors) show a relatively high degree of intra-modular connectivity and a low degree of inter-modular connectivity. High degree nodes can be classified into (i) local hubs (blue) that have a high degree centrality and low participation coefficient; and (ii) connector hubs (red) that have high degree and connect to nodes in other modules. Nodes with high betweenness centrality are located on shortest paths between nodes and can play an important role in linking different nodes, even if they have low degree (e.g., the green node supports communication between the yellow and orange modules).

(Henderson and Robinson, 2014; Roberts et al., 2016; Betzel and Bassett, 2017). Whereas, generative network models based on simple geometric rules reproduce a range of statistical properties of brain networks (Ercsey-Ravasz, 2013; Henderson and Robinson, 2014; Song et al., 2014), the spatial location of hub regions cannot be explained by geometry alone (Roberts et al., 2016), suggesting an additional role for non-geometric factors in shaping the specific topology and topography of the connectome. In this context, genes may make an important contribution to shaping complex properties, such as rich-club organization. We now turn our attention to recent studies investigating the transcriptional correlates of hub connectivity by integrating connectomic data with spatially comprehensive gene expression databases across different species and scales.

4. THE MOLECULAR CORRELATES OF HUB CONNECTIVITY

The first study to link transcriptional measures to the hub connectivity (Rubinov et al., 2015) combined gene expression data from the AMBA (Lein et al., 2007) with a mouse connectome inferred statistically from 461 tract-tracing studies (Oh et al., 2014). Data from these anterograde tracer injections into the right hemisphere were aggregated into a directed

and weighted connectivity matrix comprising of 112 bilaterally symmetrical cortical and subcortical nodes defining edge weights as normalized connection densities and ranging over four orders of magnitude, with 53% of all possible pairs of regions showing some level of non-zero connectivity. The authors identified a subset of nodes with high degree and a high participation coefficient, indicating that they were highly connected while also being connected to nodes in diverse functional systems. Using partial least squares (PLS) (Hervé, 2010), they were able to derive a linear combination of genes whose expression levels explained 48% of the variance in nodal participation coefficient. The analysis focused on a subset of 3,380 genes form the AMBA that passed quality control criteria and were assayed in at least one additional independent experiment allowing the authors to evaluate gene expression reproducibility. The genes weighting strongly on the participation-related component were enriched for GO categories, such as learning, cognition, and memory, suggesting a link between the expression of genes related to regional variations in network participation and those implicated in cognition.

In a subsequent analysis of the Allen Institute mouse connectome, Fulcher and Fornito (2016) used a parcellation comprising 213 regions linked by 3,063 connections (6.9% of all possible links), focusing on the right hemisphere only (where complete information on afferent and efferent connectivity was available), in combination with ISH measures
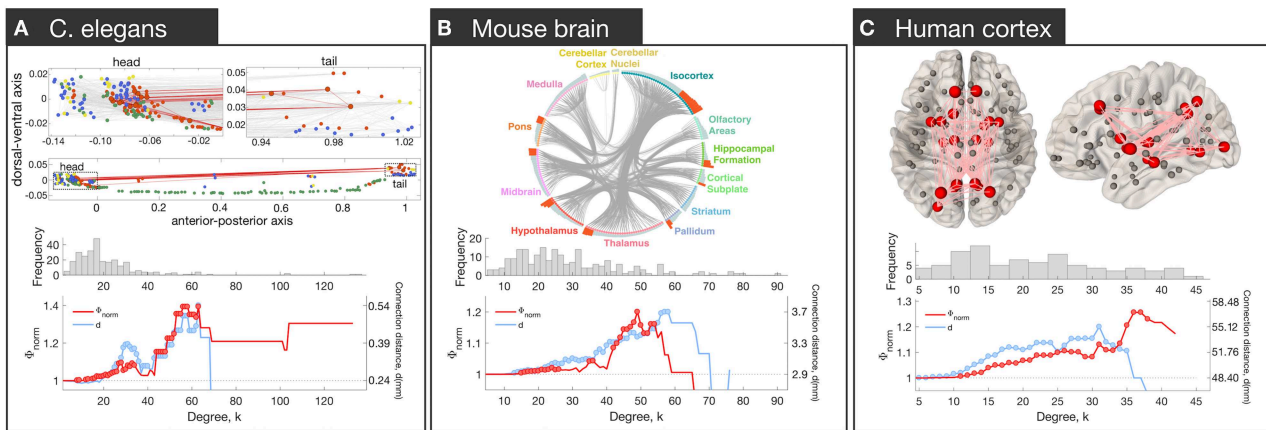
FIGURE 2 | Rich club connectivity in different species. Top row: The spatial location of hubs in *C. elegans* **(A)**, mouse **(B)**, and human **(C)**. **(A)** Neurons are represented as nodes with colors corresponding to neuron type: interneurons (red), motor neurons (green), sensory neurons (blue), multimodal neurons (yellow). Hub neurons (neurons with node degree, denoted k, > 44) are shown as circles outlined in black. Connections between hubs are shown in red; other connections shown in gray in the upper plots. The upper part represents zoomed-in plots of the head and tail that are shown as dotted rectangles in the lower plot [adapted and reproduced from (Arnatkevičiūtė et al., 2018)]. **(B)** Meso-scale connectome of the mouse. Hub regions (regions with $k > 44$) are distributed across the whole brain and contain areas in isocortex, striatum, hippocampal formation, pallidum, thalamus, hypothalamus, midbrain, pons, and cortical subplate [adapted and reproduced from (Fulcher and Fornito, 2016)]. **(C)** Macro-scale connectome of the human brain. Hub regions (regions with $k > 30$) are shown as big red spheres while other regions as smaller gray spheres. Connections between hubs are shown in pink. Hubs are bilateral: lingual gyrus, precuneus, superior frontal gyrus, superior parietal gyrus, insula, thalamus, putamen and hippocampus; right pallidum; left caudate and lateral occipital gyrus. Middle row: Distribution of degree values across nodes. In each network, the distribution is heavy-tailed, consistent with the presence of highly connected hub nodes. Bottom row: Normalized rich-club coefficient Φ_{norm} (red) and average connection distance of hub-hub links, d (blue), as a function of degree (k) at which hubs are defined. The coefficient Φ_{norm} is defined by thresholding the network at a given level of k, calculating the density of connections between hub nodes (all nodes with degree $> k$), and normalizing this value by the corresponding value obtained in an ensemble of appropriately matched surrogate graphs. The normalized coefficient therefore quantifies the degree to which the density of connections between hubs exceeds chance expectations. Since the threshold to define hubs is arbitrary, the coefficient is evaluated across all possible values of k. A rise in Φ_{norm} at high levels of k is consistent with rich-club organization. Red circles indicate Φ_{norm} values that are significantly higher than an ensemble of 10,000 null networks (permutation test $p < 0.05$). Blue circles indicate where the mean connection distance between hubs is significantly greater relative to other links in the network (one-sided Welch's t-test; $p < 0.05$).

of expression across $17,642$ genes in the AMBA (Lein et al., 2007). Their primary aim was to characterize how coupled patterns of gene expression between regions (i.e., correlated gene expression or CGE) relate to network topology. After confirming that the right hemisphere of the mouse connectome did indeed show evidence of rich-club organization, and that connections between hubs were both the most costly (measured by connection distance, reciprocity and weight) and central (measured using edge betweenness centrality and an alternative measure called communicability, that does rely on shortest path communication) connections of the network, they distinguished between three topological classes of connections following the work of van den Heuvel et al. (2012): (i) rich links, which connect two hubs (where hub is defined based on degree); (ii) feeder links, which connect a hub to non-hub (feed-out) or a non-hub to a hub (feed-in); and (iii) peripheral links, which connect two non-hubs (**Figure 3A**). Across a wide range of thresholds for defining a hub, CGE was highest for rich links, followed by feeder, and lowest for peripheral edges, with CGE showing a sharp rise at a hub threshold range that coincided with a regime in which a significant topological rich-club was observed (**Figure 3B**). This tightly coupled transcriptional activity between hub nodes defied a general trend in the brain where CGE between two areas decayed sharply (exponentially) as a function of their

distance. That is, despite connected hubs being separated by longer anatomical distances than other pairs of regions, they showed the highest levels of transcriptional coupling (note that CGE measures were corrected for this dependence). Enrichment analysis showed that this effect was driven by genes regulating the oxidative synthesis and metabolism of ATP—the primary energetic source of neuronal communication. By comparison, an enrichment analysis comparing connected to unconnected regions (regardless of whether those connections involved hubs) found significant involvement of a large number of GO categories related to synaptic plasticity and communication, axon structure, and metabolism. These findings suggest that while genes involved in forming and maintaining synapses and axons are important for establishing a connection between two regions, the primary genomic distinction between different topological classes of connections (as defined in relation to hubs) is related to the metabolic requirements of those connections.

More recently, we found a qualitatively similar pattern of elevated CGE in rich links in the nematode *C. elegans* connectome (Arnatkevičiūtė et al., 2018). Combining electron micrograph data defining the electrochemical connectome of 279 neurons (Varshney et al., 2011) with binary gene expression profiles across 948 genes (**Figure 3C**) acquired from WormBase (Harris et al., 2010), we identified the same trend for CGE to be
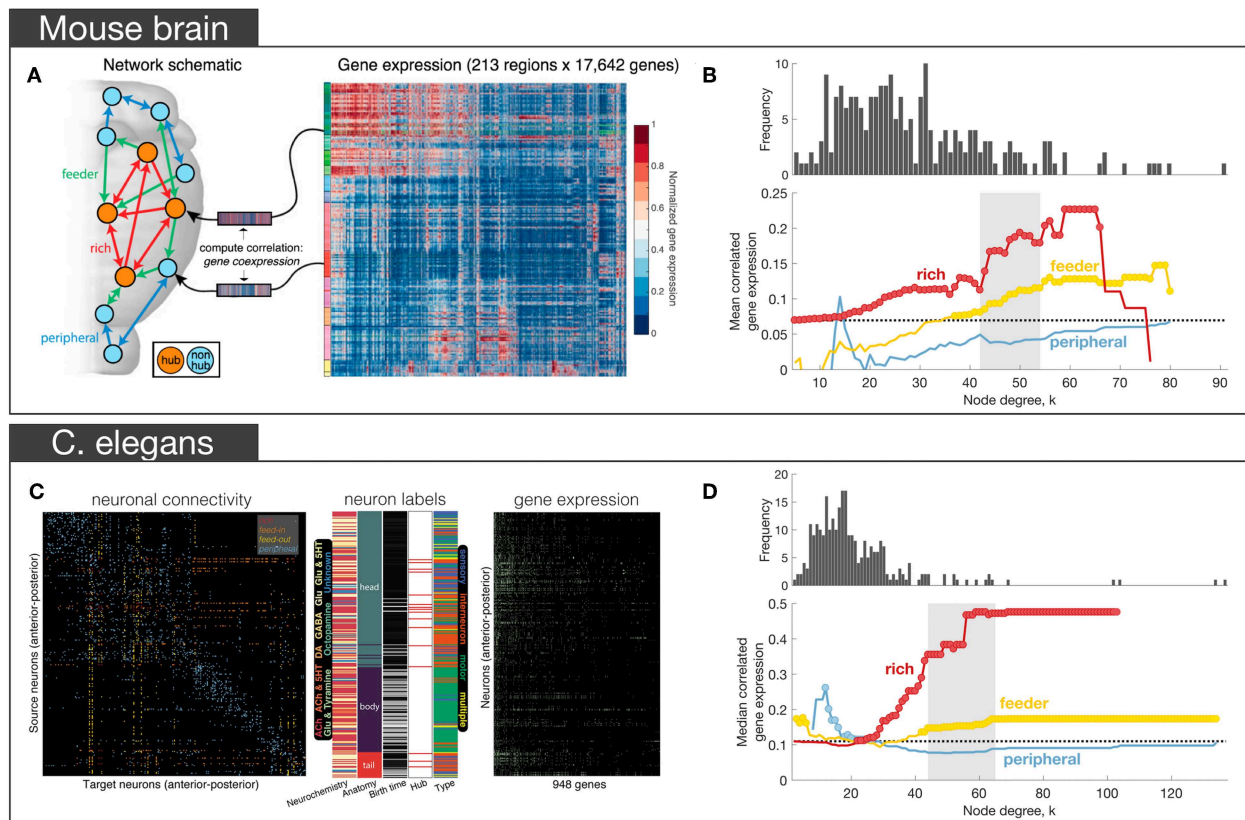
FIGURE 3 | Empirical studies investigating the transcriptional properties of hub connectivity in mouse **(A,B)** and *C. elegans* **(C,D)**. **(A)** The schematic representation of different types of connections in the mouse brain: rich (connecting a hub to a hub)—red, feeder (connecting a hub to a non-hub or a non-hub to a hub)—green, peripheral (connecting a non-hub to a non-hub)—blue. Links in the connectome were categorized across this scheme. For each region, a vector of gene expression values was extracted as the corresponding row of the region in the full gene expression matrix comprising the AMBA. The matrix represents the normalized gene expression of 17,642 genes (columns) across 213 regions (rows). Gene expression profiles for each region were then used to estimate correlated gene expression (CGE) between region pairs. **(B)** Mean correlated gene expression for rich, feeder, and peripheral links as a function of node degree (*k*) where hubs are nodes with degree > *k*. The mean CGE of rich links increases at levels of *k* that coincide with a regime where evidence of topological rich-club organization is found indicating that CGE is highest for connected pairs of network hubs. The topological rich-club regime (determined from the network topology, see **Figure 2A**) shaded gray. Circles indicate a statistically significant increase in correlated gene expression for a given link type relative to the rest of the network (one-sided Welch-s *t*-test; *p* < 0.05) [adapted and reproduced from (Fulcher and Fornito, 2016)]; **(C)** Neuron-and-synapse connectome of *C. elegans*, reconstructed for 279 neurons using electron microscopy. Connections colored according to how they connect hubs (neurons with degree > 44) and non-hubs (neurons with degree ≤ 44): red (rich links connecting hubs), orange (feed-in links connecting a non-hub to a hub), yellow (feed-out links connecting a hub to a non-hub), blue (peripheral links connecting non-hubs). Middle: additional data acquired for each neuron, such as its: chemically secreted transmitter, anatomical location, birth time, hub status and neuronal type. Right binary gene expression profile for each of the 279 neurons (rows) across 948 genes (columns). **(D)** Median CGE for each connection type (feed-in and feed-out connections are combined and represented as feeder) as a function of node degree *k*. The topological rich-club regime (determined from the network topology, see **Figure 2A**) shaded gray. Circles indicate a statistically significant increase in CGE in a given link type relative to the rest of the network (one-sided Wilcoxon rank sum test, *p* < 0.05) [adapted and reproduced from (Arnatkevičiūtė et al., 2018)].

highest for rich links, followed by feeder, and then peripheral edges (**Figure 3D**). The involvement of metabolic genes in rich-club connectivity—as in the mesoscopic mouse connectome (Fulcher and Fornito, 2016)—could not be confirmed due to limited gene expression data in the worm, but analysis of the available data indicated that glutamate signaling and neuronal communication genes made the strongest contribution to elevated CGE for hub-hub connections (Arnatkevičiūtė et al., 2018). Leveraging the extensive additional data on neuronal phenotypes available for the worm, we found that elevated CGE for connected hubs could not be explained by a range

of other properties, such as neuronal lineage distance (number of cell divisions separating pairs of neurons from a common ancestor), differences in birth time, neuronal subtype (sensory, motor, or interneuron), chemically secreted neurotransmitter, anatomical separation distance or topological module affiliation. However, the effect did seem to be driven by the fact that most hubs in the worm connectome are command interneurons, a specialized class of neurons that regulates motion. Motion is one of the more complex behaviors in the worm's repertoire, and these findings parallel evidence in primates that network hubs are primarily located in association cortices, which are

thought to mediate higher-order cognition (Achard et al., 2006; Sporns et al., 2007). Thus, despite numerous differences in the data, including different gene annotation methods (~ 20 000 ISH genes in mouse vs. ~ 1, 000 binary literature-curated annotations in worm), the type of the neural system (spatially continuous macroscopic mouse brain *vs* spatially separated *C. elegans* nervous system), and the orders of magnitude differences in scale, both studies demonstrated the same general pattern of increased transcriptional similarity across topologically central hub nodes.

In light of the findings in both mouse and *C. elegans*, where several groups of genes implicated in cognition (Rubinov et al., 2015), oxidative metabolism (Fulcher and Fornito, 2016), and neuronal communication (Arnatkevičiūtė et al., 2018) have been identified as being related to hub connectivity, one could wonder whether the same genes are involved in the hub connectivity of the human brain. The first analysis to link gene expression and hub connectivity in humans was performed by Vértes et al. (2016), who combined resting-state fMRI (rs-fMRI) data with the high coverage genome-wide gene expression from AHBA (Hawrylycz et al., 2012). Rendering rs-fMRI data for 285 cortical regions as a binary undirected network, thresholded to retain 10% of all possible connections, they measured three different properties of each node: its within-module connectivity, its participation coefficient (between-module connectivity), and its average Euclidean distance from other nodes. PLS identified three components that collectively accounted for 37% of the total variance in nodal metrics

with the first component exhibiting a positive correlation with intra-modular degree and a negative correlation with average nodal distance, corresponding to high degree nodes that mostly form short-range within-module connections. Genes positively loading on this component were enriched for GO categories related to transcriptional regulation. The second component was positively related to both the participation coefficient and average nodal distance, thus representing nodes with long connections that extend between modules, consistent with the integrative hubs of the network (**Figure 4A**). As seen in the analysis of the structural connectivity analysis of the mouse (Fulcher and Fornito, 2016), genes loading positively on this component were enriched in GO categories related to oxidative metabolism and mitochondrial function. These genes also showed significant over-representation for a set of 19 genes (Krienen et al., 2016) selectively enriched in the supragranular layers of the human cortex (HSE-human supragranular enriched genes) with some of those genes being implicated in the formation of corticocortical projections emanating from the higher layers of the cortex (Krienen et al., 2016). Together these findings suggest that hubs across species demonstrate conserved transcriptional properties related to their high metabolic demands.

It is well-known that the human brain undergoes an extended period of development during adolescence that is critical for brain maturation and coincides with the period of peak risk for many mental disorders (Paus et al., 2008). Some of those developmental changes particularly target hub

FIGURE 4 | Empirical studies investigating the transcriptional properties of hub connectivity in human. **(A)** A schematic representation of the modular organization of the connectome demonstrating the key properties of inter- and intra- modular hubs based on Vértes et al. (2016). Intra-modular hubs (blue nodes) mostly connect nodes within the same module and have relatively short connection distances; characterized by the PLS1. Intra-modular hubs (red nodes) have a more diverse connectivity profile with connections extending long distances and connecting nodes from different modules; characterized by the PLS2. Size and color saturation of the nodes in the connectome corresponds to the regional scores on PLS1 (Intra-modular hub) and PLS2 (Inter-modular hub) to represent the spatial pattern of transcriptional profiles [adapted and modified from (Vértes et al., 2016)]. **(B)** Gene expression and cortical consolidation in adolescence based on Whitaker et al. (2016), (top) spatial topography of the second component from a PLS analysis corresponding to cortical consolidation during adolescence, defined as cortical shrinkage/myelination. Genes identified in this profile are related to synaptic transmission and risk to schizophrenia, among others, and are overexpressed in prefrontal areas of the cortex; (bottom) hubs in the structural covariance network experience faster rates of cortical thinning (CT) and myelination. The PLS2 gene expression profile is also significantly associated with degree, meaning that hubs are likely to over-express those genes [adapted and modified from (Whitaker et al., 2016)].

regions (Dennis et al., 2013; Hwang et al., 2013; Baker et al., 2015; for a review see Cao et al., 2016). Whitaker et al. (2016) examined a large sample of adolescents (279, aged 14–24 years old) and found that topologically central hubs of the cortical structural covariance networks undergo an increased rate of consolidation, defined by increased cortical thinning and enhanced myelination (**Figure 4B**). Components of transcriptional variance that correlated with this consolidation were extracted using PLS, employing the full set of 20 737 genes from the AHBA. The first two components explaining 28% of the variance in MRI measures were related to the baseline measures of cortical thickness and myelination (PLS1), and cortical shrinkage and myelination—consolidation over time (PLS2) (**Figure 4B**), respectively. The PLS2 component involved contributions from genes regulating synaptic transmission and a set of genes linked to risk for schizophrenia, suggesting that deviation from the normal developmental consolidation of hub regions might manifest as an intermediate phenotype for schizophrenia (Whitaker et al., 2016), consistent with evidence that hubs are disproportionately impacted by the disease (van den Heuvel et al., 2013b; Crossley et al., 2014; Klauser et al., 2016) and that regional variations in the expression of schizophrenia risk genes track the regional variations in the magnitude of group differences in connectivity between controls and patients (Romme et al., 2017).

Importantly, this work implies that genes involved in the development of hubs, which relate to myelination and synaptic transmission, are distinct from those implicated in cross-sectional studies of adult hub connectivity, which implicate metabolic genes. In other words, the genetic mechanisms underlying the development of hub connectivity may differ from those involved in sustaining the functional role that hubs play in a mature neuronal system. The further development of brain-wide atlases of developmental changes in gene expression will help shed light on how such differences can be leveraged to gain insight into the development of different brain disorders.

5. CONCLUSIONS AND FURTHER DIRECTIONS

Brain-wide gene expression atlases provide exciting opportunities to link different scales of brain organization. At the same time, integrating such data with connectomic measures poses challenges. Given the nascence of this field, no standardized data processing pipelines have been developed, with widespread inconsistencies in processing of the same transcriptional data across studies (Arnatkevičiūtė et al., 2019) complicating direct comparison between findings, even within the same species.

Nonetheless, the available studies—conducted in diverse species and using different measures of brain connectivity and gene expression acquired at different resolution scales—point to a conserved transcriptional signature of hub connectivity related to genes regulating neuronal communication and metabolism, consistent with the high centrality and metabolic cost of hub regions (Bullmore and Sporns, 2009).

One limitation affecting the human data is that the gene expression measures are derived from bulk tissue samples. The cellular composition of these samples can influence measured gene expression patterns, such that two samples can differ in their transcriptional properties simply due to the differences in the density of distinct cell types. Single-cell transcriptomics is able to provide precise gene expression measurements in individual cells, thus resolving cell-specific transcriptional profiles. While scRNA-seq is not currently feasible for the whole human brain, the expression profiles of specific cell groups in the adult (Johnson et al., 2015; Hu and Wang, 2017; Picardi et al., 2017) and developing brain (Zhong et al., 2018) are being characterized.

These limitations notwithstanding, the consistency of results considered here—often identified through unbiased, data-driven techniques—demonstrate the potential utility of brain-wide transcriptomic measures in yielding biologically meaningful insights to otherwise abstract graph-theoretical structures, such as hubs and other neural phenotypes. With the availability of new resources and developments in neuroimaging, the combination of such data across resolution scales offers a promising way forward for uncovering the molecular mechanisms that drive the large-scale organization of the connectome.

AUTHOR CONTRIBUTIONS

AA wrote and edited the manuscript. BF and AF provided feedback, structured, and edited the manuscript. All the authors planned the structure of the manuscript.

FUNDING

ACKNOWLEDGMENTS

REFERENCES

Achard, S., Salvador, R., Whitcher, B., Suckling, J., and Bullmore, E. (2006). A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs. *J. Neurosci.* 26, 63–72. doi: 10.1523/JNEUROSCI.3874-05.2006

Albert, R., Jeong, H., and Barabási, A. L. (2000). Error and attack tolerance of complex networks. *Nature* 406, 378–382. doi: 10.1038/35019019

Anderson, K. M., Krienen, F. M., Choi, E. Y., Reinen, J. M., Yeo, B. T. T., and Holmes, A. J. (2018). Gene expression links functional networks across cortex and striatum. *Nat. Commun.* 9:1428. 1428. doi: 10.1038/s41467-018-03811-x

Arnatkevičiūtė, A., Fulcher, B. D., and Fornito, A. (2019). A practical guide to linking brain-wide gene expression and neuroimaging data. *Nauroimage* 189, 353-367. doi: 10.1016/j.neuroimage.2019.01.011

Arnatkevičiūtė, A., Fulcher, B. D., Pocock, R., and Fornito, A. (2018). Hub connectivity, neuronal diversity, and gene expression in the Caenorhabditis elegans connectome. *PLoS Comput. Biol.* 14:e1005989. doi: 10.1371/journal.pcbi.1005989

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Davis, A. P., et al. (2000). Gene ontology: tool for the unification of biology. *Nat. Genet.* 25, 25–29. doi: 10.1038/75556

Ayoub, A. E., Oh, S., Xie, Y., Leng, J., Cotney, J., Dominguez, M. H., et al. (2011). Transcriptional programs in transient embryonic zones of the cerebral cortex defined by high-resolution mRNA sequencing. *Proc. Natl. Acad. Sci. U.S.A.* 108, 14950–14955. doi: 10.1073/pnas.1112213108

Baker, S. T. E., Lubman, D. I., Yücel, M., Allen, N. B., Whittle, S., Fulcher, B. D., et al. (2015). Developmental changes in brain network hub connectivity in late adolescence. *J. Neurosci.* 35, 9078–9087. doi: 10.1523/jneurosci.5043-14.2015

Bakken, T. E., Miller, J. A., Ding, S. L., Sunkin, S. M., Smith, K. A., Ng, L., et al. (2016). A comprehensive transcriptional map of primate brain development. *Nature* 535, 367–375. doi: 10.1038/nature18637

Baruch, L., Itzkovitz, S., Golan-Mashiach, M., Shapiro, E., and Segal, E. (2008). Using expression profiles of *Caenorhabditis elegans* neurons to identify genes that mediate synaptic connectivity. *PLoS Comput. Biol.* 4, 1–10. doi: 10.1371/journal.pcbi.1000120

Bassett, D. S., and Bullmore, E. T. (2009). Human brain networks in health and disease. *Curr. Opin. Neurol.* 22, 340–347. doi: 10.1097/WCO.0b013e32832d93dd

Bassett, D. S., and Bullmore, E. T. (2017). Small-world brain networks revisited. *Neuroscience* 23, 499–516. doi: 10.1177/1073858416667720

Bassett, D. S., Greenfield, D. L., Meyer-Lindenberg, A., Weinberger, D. R., Moore, S. W., and Bullmore, E. T. (2010). Efficient physical embedding of topologically complex information processing networks in brains and computer circuits. *PLoS Comput. Biol.* 6:e1000748. doi: 10.1371/journal.pcbi.1000748

Belgard, T., Marques, A., Oliver, P., Abaan, H. O., Sirey, T., Hoerder-Suabedissen, A., et al. (2011). A transcriptomic atlas of mouse neocortical layers. *Neuron* 71, 605–616. doi: 10.1016/j.neuron.2011.06.039

Berchtold, N. C., Cribbs, D. H., Coleman, P. D., Rogers, J., Head, E., Kim, R., et al. (2008). Gene expression changes in the course of normal brain aging are sexually dimorphic. *Proc. Natl. Acad. Sci. U.S.A.* 105, 15605–15610. doi: 10.1073/pnas.0806883105

Betzel, R. F., and Bassett, D. S. (2017). Multi-scale brain networks. *Neuroimage* 160, 73–83. doi: 10.1016/J.NEUROIMAGE.2016.11.006

Blondel, V. D., Guillaume, J. L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp.* 2008:P10008. doi: 10.1088/1742-5468/2008/10/P10008

Bohlken, M. M., Mandl, R. C., Brouwer, R. M., van den Heuvel, M. P., Hedman, A. M., Kahn, R. S., et al. (2014). Heritability of structural brain network topology: a DTI study of 156 twins. *Hum. Brain Mapp.* 35, 5295–5305. doi: 10.1002/hbm.22550

Brown, J. A., Terashima, K. H., Burggren, A. C., Ercoli, L. M., Miller, K. J., Small, G. W., et al. (2011). Brain network local interconnectivity loss in aging APOE-4 allele carriers. *Proc. Natl. Acad. Sci. U.S.A.* 108, 20760–20765. doi: 10.1073/pnas.1109038108

Bullmore, E., and Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat. Rev. Neurosci.* 10, 186–198. doi: 10.1038/nrn2575

Bullmore, E., and Sporns, O. (2012). The economy of brain network organization. *Nat. Rev. Neurosci.* 13, 336–49. doi: 10.1038/nrn3214

Burt, J. B., Demirtas, M., Eckner, W. J., Navejar, N. M., Ji, J. L., Martin, W. J., et al. (2017). Hierarchy of transcriptomic specialization across human cortex captured by myelin map topography. *Nat. Neurosci.* 21, 1251–1259. doi: 10.1101/199703

Buzsáki, G., and Draguhn, A. (2004). Neuronal oscillations in cortical networks. *Science* 304, 1926–1929. doi: 10.1126/science.1099745

Cao, M., Huang, H., Peng, Y., Dong, Q., and He, Y. (2016). Toward developmental connectomics of the human brain. *Front. Neuroanat.* 10:25. doi: 10.3389/fnana.2016.00025

Cioli, C., Abdi, H., Beaton, D., Burnod, Y., and Mesmoudi, S. (2014). Differences in human cortical gene expression match the temporal properties of large-scale functional networks. *PLoS ONE* 9:e115913. doi: 10.1371/journal.pone.0115913

Colantuoni, C., Lipska, B. K., Ye, T., Hyde, T. M., Tao, R., Leek, J. T., et al. (2011). Temporal dynamics and genetic control of transcription in the human prefrontal cortex. *Nature* 478, 519–23. doi: 10.1038/nature10524

Colclough, G. L., Smith, S. M., Nichols, T. E., Winkler, A. M., Sotiropoulos, S. N., Glasser, M. F., et al. (2017). The heritability of multi-modal connectivity in human brain activity. *Elife* 6:e20178. doi: 10.7554/eLife.20178

Colizza, V., Flammini, A., Serrano, M. A., and Vespignani, A. (2006). Detecting rich-club ordering in complex networks. *Nat. Phys.* 2, 110–115. doi: 10.1038/nphys209

Crossley, N. A., Mechelli, A., Scott, J., Carletti, F., Fox, P. T., McGuire, P., et al. (2014). The hubs of the human connectome are generally implicated in the anatomy of brain disorders. *Brain J. Neurol.* 137, 2382–95. doi: 10.1093/brain/awu132

Darmanis, S., Sloan, S. A., Zhang, Y., Enge, M., Caneda, C., Shuer, L. M., et al. (2015). A survey of human brain transcriptome diversity at the single cell level. *Proc. Natl. Acad. Sci. U.S.A.* 112, 7285–7290. doi: 10.1073/pnas.1507125112

de Reus, M. A., and van den Heuvel, M. P. (2013). Rich club organization and intermodule communication in the cat connectome. *J. Neurosci.* 33, 12929–12939. doi: 10.1523/JNEUROSCI.1448-13.2013

de Reus, M. A., and van den Heuvel, M. P. (2014). Simulated rich club lesioning in brain networks: a scaffold for communication and integration? *Front. Hum. Neurosci.* 8:647. doi: 10.3389/fnhum.2014.00647

Dennis, E. L., Jahanshad, N., Rudie, J. D., Brown, J. A., Johnson, K., McMahon, K. L., et al. (2011). Altered structural brain connectivity in healthy carriers of the autism risk gene, CNTNAP2. *Brain Connect.* 1, 447–459. doi: 10.1089/brain.2011.0064

Dennis, E. L., Jahanshad, N., Toga, A. W., McMahon, K. L., de Zubicaray, G. I., Hickie, I., et al. (2013). Development of the rich club in brain connectivity networks from 438 adolescents & adults aged 12 to 30. *Proc. IEEE Int. Symp. Biomed. Imaging* 2013, 624–627 624–627 doi: 10.1109/ISBI.2013.6556552

Eising, E., Huisman, S. M., Mahfouz, A., Vijfhuizen, L. S., Anttila, V., Winsvold, B. S., et al. (2016). Gene co-expression analysis identifies brain regions and cell types involved in migraine pathophysiology: a GWAS-based study using the Allen Human Brain Atlas. *Hum. Genet.* 135, 425–439. doi: 10.1007/s00439-016-1638-x

Ercsey-Ravasz, A. M., Markov, N. T., Lamy, C.,Van Essen, D. C., Knoblauch, K., Toroczkai, Z., et al. (2013). A predictive network model of cerebral cortical connectivity based on a distance rule. *Neuron* 80, 184–197. doi: 10.1016/j.neuron.2013.07.036

Fakhry, A., and Ji, S. (2015). High-resolution prediction of mouse brain connectivity using gene expression patterns. *Methods* 73, 71–78. doi: 10.1016/j.ymeth.2014.07.011

Fakhry, A., Zeng, T., Peng, H., and Ji, S. (2015). Global analysis of gene expression and projection target correlations in the mouse brain. *Brain Inform.* 2, 107–117. doi: 10.1007/s40708-015-0014-2

Felleman, D. J., and Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex* 1, 1–47.

Fertuzinhos, S., Li, M., Kawasawa, Y., Ivic, V., Franjic, D., Singh, D., et al. (2014). Laminar and temporal expression dynamics of coding and noncoding RNAs in the mouse neocortex. *Cell Rep.* 6, 938–950. doi: 10.1016/j.celrep.2014.01.036

Forest, M., Iturria-Medina, Y., Goldman, J. S., Kleinman, C. L., Lovato, A., Oros Klein, K., et al. (2017). Gene networks show associations with seed region connectivity. *Hum. Brain Mapp.* 38, 3126–3140. doi: 10.1002/hbm.23579

Fornito, A., Arnatkevičiūtė, A., and Fulcher, B. D. (2019). Bridging the gap between connectome and transcriptome. *Trends Cogn. Sci.* 23, 34–50. doi: 10.1016/j.tics.2018.10.005

Fornito, A., Zalesky, A., Bassett, D. S., Meunier, D., Ellison-Wright, I., Yücel M, et al. (2011). Genetic influences on cost-efficient organization of human cortical functional networks. *J. Neurosci.* 31, 3261–3270. doi: 10.1523/JNEUROSCI.4858-10.2011

Fornito, A., Zalesky, A., and Breakspear, M. (2013). Graph analysis of the human connectome: promise, progress, and pitfalls. *Neuroimage* 80, 426–444. doi: 10.1016/j.neuroimage.2013.04.087

Fornito, A., Zalesky, A., and Breakspear, M. (2015). The connectomics of brain disorders. *Nat. Rev. Neurosci.* 16, 159–172. doi: 10.1038/nrn3901

Fornito, A., Zalesky, A., and Bullmore, E. (2016). *Fundamentals of Brain Network Analysis*. San Diego, CA: Academic Press.

Fraser, H. B., Khaitovich, P., Plotkin, J. B., Pääbo, S., and Eisen, M. B. (2005). Aging and gene expression in the primate brain. *PLoS Biol.* 3:e274. doi: 10.1371/journal.pbio.0030274

French, L., and Pavlidis, P. (2011). Relationships between gene expression and brain wiring in the adult rodent brain. *PLoS Comput. Biol.* 7:e1001049. doi: 10.1371/journal.pcbi.1001049

Friston, K. J. (1994). Functional and effective connectivity in neuroimaging: a synthesis. *Hum. Brain Mapp.* 2, 56–78. doi: 10.1002/hbm.460020107

Fulcher, B. D., and Fornito, A. (2016). A transcriptional signature of hub connectivity in the mouse connectome. *Proc. Natl. Acad. Sci. U.S.A.* 113, 1435–1440. doi: 10.1073/pnas.1513302113

Fürth, D., Vaissiére, T., Tzortzi, O., Xuan, Y., Märtin, A., Lazaridis, I., et al. (2018). An interactive framework for whole-brain maps at cellular resolution. *Nat. Neurosci.* 21, 139–149. doi: 10.1038/s41593-017-0027-7

Gamanut, R., Kennedy, H., Toroczkai, Z., Ercsey-Ravasz, M., Van Essen, D. C., Knoblauch, K., et al. (2018). The mouse cortical connectome, characterized by an ultra-dense cortical graph, maintains specificity by distinct connectivity profiles. *Neuron* 97, 698–715.e10. doi: 10.1016/J.NEURON.2017.12.037

Goel, P., Kuceyeski, A., LoCastro, E., and Raj, A. (2014). Spatial patterns of genome-wide expression profiles reflect anatomic and fiber connectivity architecture of healthy human brain. *Hum. Brain Mapp.* 35, 4204–4218. doi: 10.1002/hbm.22471

Goñi, J., van den Heuvel, M. P., Avena-Koenigsberger, A., Velez de Mendizabal, N., Betzel, R. F., Griffa, A., et al. (2014). Resting-brain functional connectivity predicted by analytic measures of network communication. *Proc. Natl. Acad. Sci. U.S.A.* 111, 833–838. doi: 10.1073/pnas.1315529111

Guimerá, R., and Amaral, L. A. N. (2005). Functional cartography of complex metabolic networks. *Nature* 433, 895–900. doi: 10.1038/nature03288

Harriger, L., van den Heuvel, M. P., and Sporns, O. (2012). Rich club organization of macaque cerebral cortex and its role in network communication. *PLoS ONE* 7:e46497. doi: 10.1371/journal.pone.0046497

Harris, T. W., Antoshechkin, I., Bieri, T., Blasiar, D., Chan, J., Chen, W. J., et al. (2010). WormBase: a comprehensive resource for nematode research. *Nucleic Acids Res.* 38, D463–D467. doi: 10.1093/nar/gkp952

Hawrylycz, M., Miller, J. A., Menon, V., Feng, D., Dolbeare, T., Guillozet-Bongaarts, A. L., et al. (2015). Canonical genetic signatures of the adult human brain. *Nat. Neurosci.* 18, 1832–1844. doi: 10.1038/nn.4171

Hawrylycz, M. J., Lein, E. S., Guillozet-Bongaarts, A. L., Shen, E. H., Ng, L., Miller, J. A., et al. (2012). An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* 489, 391–399. doi: 10.1038/nature11405

Henderson, J. A., and Robinson, P. A. (2014). Relations between the geometry of cortical gyrification and white-matter network architecture. *Brain Connect.* 4, 112–130. doi: 10.1089/brain.2013.0183

Hervé, A. (2010). Partial least squares regression and projection on latent structure regression (PLS Regression). *Wiley Interdiscip. Rev. Comput. Stat.* 2, 97–106. doi: 10.1002/wics.51

Hu, G., and Wang, G. Z. (2017). Decoding nervous system by single-cell RNA sequencing. *Quant. Biol.* 5, 210–214. doi: 10.1007/s40484-017-0116-3

Hwang, K., Hallquist, M. N., and Luna, B. (2013). The development of hub architecture in the human functional brain network. *Cereb. Cortex* 23, 2380–2393. doi: 10.1093/cercor/bhs227

Irizarry, R. A., Wang, C., Zhou, Y., and Speed, T. P. (2009). Gene set enrichment analysis made simple. *Stat. Methods Med. Res.* 18, 565–575. doi: 10.1177/0962280209351908

Jahanshad, N., Rajagopalan, P., Hua, X., Hibar, D. P., Nir, T. M., Toga, A. W., et al. (2013). Genome-wide scan of healthy human connectome discovers SPON1 gene variant influencing dementia severity. *Proc. Natl. Acad. Sci. U.S.A.* 110, 4768–4773. doi: 10.1073/pnas.1216206110

Ji, S., Fakhry, A., and Deng, H. (2014). Integrative analysis of the connectivity and gene expression atlases in the mouse brain. *Neuroimage* 84, 245–253. doi: 10.1016/j.neuroimage.2013.08.049

Johnson, G. A., Badea, A., Brandenburg, J., Cofer, G., Fubara, B., Liu, S., et al. (2010). Waxholm space: an image-based reference for coordinating mouse brain research. *Neuroimage* 53, 365–372. doi: 10.1016/j.neuroimage.2010.06.067

Johnson, M. B., Wang, P. P., Atabay, K. D., Murphy, E. A., Doan, R. N., Hecht, J. L., et al. (2015). Single-cell analysis reveals transcriptional heterogeneity of neural progenitors in human cortex. *Nat. Neurosci.* 18, 637–646. doi: 10.1038/nn.3980

Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30. doi: 10.1093/nar/28.1.27

Kang, H. J., Kawasawa, Y. I., Cheng, F., Zhu, Y., Xu, X., Li, M., et al. (2011). Spatio-temporal transcriptome of the human brain. *Nature* 478, 483–489. doi: 10.1038/nature10523

Kaufman, A., Dror, G., Meilijson, I., and Ruppin, E. (2006). Gene expression of Caenorhabditis elegans neurons carries information on their synaptic connectivity. *PLoS Comput. Biol.* 2:e167. doi: 10.1371/journal.pcbi.0020167

Keil, J. M., Qalieh, A., and Kwan, K. Y. (2018). Brain transcriptome databases: a user's guide. *J. Neurosci.* 10, 1930–1917. doi: 10.1523/JNEUROSCI.1930-17.2018

Keo, A., Aziz, N. A., Dzyubachyk, O., van der Grond, J., van Roon-Mom, W. M. C., Lelieveldt, B. P. F., et al. (2017). Co-expression patterns between ATN1 and ATXN2 coincide with brain regions affected in Huntington's disease. *Front. Mol. Neurosci.* 10:399. doi: 10.3389/fnmol.2017.00399

Klauser, P., Baker, S. T., Cropley, V. L., Bousman, C., Fornito, A., and Cocchi, L. (2016). White matter disruptions in schizophrenia are spatially widespread and topologically converge on brain network hubs. *Schizophr. Bull.* 43, 425–435. doi: 10.1093/schbul/sbw100

Krienen, F. M., Yeo, B. T. T., Ge, T., Buckner, R. L., and Sherwood, C. C. (2016). Transcriptional profiles of supragranular-enriched genes associate with corticocortical network architecture in the human brain. *Proc. Natl. Acad. Sci. U.S.A.* 113, E469–E478. doi: 10.1073/pnas.1510903113

Kumar, A., Gibbs, J. R., Beilina, A., Dillman, A., Kumaran, R., Trabzuni, D., et al. (2013). Age-associated changes in gene expression in human brain and isolated neurons. *Neurobiol. Aging* 34, 1199–1209. doi: 10.1016/j.neurobiolaging.2012.10.021

Lein, E. S., Hawrylycz, M. J., Ao, N., Ayres, M., Bensinger, A., Bernard, A., et al. (2007). Genome-wide atlas of gene expression in the adult mouse brain. *Nature* 445, 168–176. doi: 10.1038/nature05453

Liang, X., Zou, Q., He, Y., and Yang, Y. (2013). Coupling of functional connectivity and regional cerebral blood flow reveals a physiological basis for network hubs of the human brain. *Proc. Natl. Acad. Sci. U.S.A.* 110, 1929–1934. doi: 10.1073/pnas.1214900110

Lichtman, J. W., and Denk, W. (2011). The big and the small: challenges of imaging the brain's circuits. *Science* 334, 618–623. doi: 10.1126/science.1209168

Liu, H., Bebu, I., and Li, X. (2010). Microarray probes and probe sets. *Front. Biosci. (Elite Ed)* 2, 325–338. doi: 10.1016/j.biotechadv.2011.08.021.Secreted

Mancarci, B. O., Toker, L., Tripathy, S. J., Li, B., Rocco, B., Sibille, E., et al. (2017). Cross-laboratory analysis of brain cell type transcriptomes with applications to interpretation of bulk tissue data. *eNeuro* 4:ENEURO.0212–17.2017. doi: 10.1523/ENEURO.0212-17.2017

Markett, S., de Reus, M. A., Reuter, M., Montag, C., Weber, B., Schoene-Bake, J. C., et al. (2017). Variation on the dopamine D2 receptor gene (DRD2) is associated with basal ganglia-to-frontal structural connectivity. *Neuroimage* 155, 473–479. doi: 10.1016/j.neuroimage.2017.04.005

Markov, N. T., Ercsey-Ravasz, M. M., Ribeiro Gomes, A. R., Lamy, C., Magrou, L., Vezoli, J., et al. (2014). A weighted and directed interareal connectivity matrix for macaque cerebral cortex. *Cereb. Cortex* 24, 17–36. doi: 10.1093/cercor/bhs270

McColgan, P., Gregory, S., Seunarine, K. K., Razi, A., Papoutsi, M., Johnson, E., et al. (2018). Brain regions showing white matter loss in Huntington's disease are enriched for synaptic and metabolic genes. *Biol. Psychiatry* 83, 456–465. doi: 10.1016/J.BIOPSYCH.2017.10.019

Mesulam, M. M. (1998). From sensation to cognition. *Brain J. Neurol.* 121, 1013–1052.

Meunier, D., Lambiotte, R., Fornito, A., Ersche, K. D., and Bullmore, E. T. (2009). Hierarchical modularity in human brain functional networks. *Front. Neuroinform.* 3:37. doi: 10.3389/neuro.11.037.2009

Meyer, K., and Damasio, A. (2009). Convergence and divergence in a neural architecture for recognition and memory. *Trends Neurosci.* 32, 376–382. doi: 10.1016/J.TINS.2009.04.002

Miller, J. A., Ding, S. L., Sunkin, S. M., Smith, K. A., Ng, L., Szafer, A., et al. (2014). Transcriptional landscape of the prenatal human brain. *Nature* 508, 199–206. doi: 10.1038/nature13185

Mišić, B., Betzel, R., Nematzadeh, A., Goñi, J., Griffa, A., Hagmann, P., et al. (2015). Cooperative and competitive spreading dynamics on the human connectome. *Neuron* 86, 1518–1529. doi: 10.1016/J.NEURON.2015.05.035

Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* 5, 621–628. doi: 10.1038/nmeth.1226

Myers, A. J., Gibbs, J. R., Webster, J. A., Rohrer, K., Zhao, A., Marlowe, L., et al. (2007). A survey of genetic human cortical gene expression. *Nat. Genet.* 39, 1494–1499. doi: 10.1038/ng.2007.16

Negi, S. K., and Guda, C. (2017). Global gene expression profiling of healthy human brain and its application in studying neurological disorders. *Sci. Rep.* 7:897. doi: 10.1038/s41598-017-00952-9

Oh, S. W., Harris, J. A., Ng, L., Winslow, B., Cain, N., Mihalas, S., et al. (2014). A mesoscale connectome of the mouse brain. *Nature* 508, 207–214. doi: 10.1038/nature13186

Okoniewski, M. J., and Miller, C. J. (2006). Hybridization interactions between probesets in short oligo microarrays lead to spurious correlations. *BMC Bioinformatics* 7:276. doi: 10.1186/1471-2105-7-276

Oldham, S., Fulcher, B., Parkes, L., Arnatkevičiūtė, A., Suo, C., and Fornito, A. (2018). Consistency and differences between centrality metrics across distinct classes of networks. Available online at: https://arxiv.org/pdf/1805.02375.pdf

Pantazatos, S. P., and Li, X. (2017). Commentary: BRAIN NETWORKS. Correlated gene expression supports synchronous activity in brain networks. *Science* 348, 1241–4. *Front. Neurosci.* 11:412. doi: 10.3389/fnins.2017.00412

Parkes, L., Fulcher, B. D., Yücel, M., and Fornito, A. (2017). Transcriptional signatures of connectomic subregions of the human striatum. *Genes Brain Behav.* 16, 647–663. doi: 10.1111/gbb.12386

Paus, T., Keshavan, M., and Giedd, J. N. (2008). Why do many psychiatric disorders emerge during adolescence? *Nat. Rev. Neurosci.* 9, 947–957. doi: 10.1038/nrn2513

Picardi, E., Horner, D. S., and Pesole, G. (2017). Single-cell transcriptomics reveals specific RNA editing signatures in the human brain. *RNA* 23, 860–865. doi: 10.1261/rna.058271.116

Poulin, J. F., Tasic, B., Hjerling-Leffler, J., Trimarchi, J. M., and Awatramani, R. (2016). Disentangling neural cell diversity using single-cell transcriptomics. *Nat. Neurosci.* 19, 1131–1141. doi: 10.1038/nn.4366

Power, J. D., Cohen, A. L., Nelson, S. M., Wig, G. S., Barnes, K. A., Church, J. A., et al. (2011). Functional network organization of the human brain. *Neuron* 72, 665–678. doi: 10.1016/j.neuron.2011.09.006

Richiardi, J., Altmann, A., Milazzo, A. C., Chang, C., Chakravarty, M. M., Banaschewski, T., et al. (2015). Correlated gene expression supports synchronous activity in brain networks. *Science* 348, 1241–1244. doi: 10.1126/science.1255905

Rittman, T., Rubinov, M., Vértes, P. E., Patel, A. X., Ginestet, C. E., Ghosh, B. C., et al. (2016). Regional expression of the MAPT gene is associated with loss of hubs in brain networks and cognitive impairment in Parkinson disease and progressive supranuclear palsy. *Neurobiol. Aging* 48, 153–160. doi: 10.1016/j.neurobiolaging.2016.09.001

Roberts, J. A., Perry, A., Lord, A. R., Roberts, G., Mitchell, P. B., Smith, R. E., et al. (2016). The contribution of geometry to the human connectome. *Neuroimage* 124, 379–393. doi: 10.1016/j.neuroimage.2015.09.009

Romero-Garcia, R., Whitaker, K. J., Váša, F., Seidlitz, J., Shinn, M., Fonagy, P., et al. (2018). Structural covariance networks are coupled to expression of genes enriched in supragranular layers of the human cortex. *Neuroimage* 171, 256–267. doi: 10.1016/J.NEUROIMAGE.2017.12.060

Romme, I. A. C., de Reus, M. A., Ophoff, R. A., Kahn, R. S., and van den Heuvel, M. P. (2017). Connectome disconnectivity and cortical gene expression in patients with schizophrenia. *Biol. Psychiatry* 81, 495–502. doi: 10.1016/j.biopsych.2016.07.012

Royce, T. E., Rozowsky, J. S., and Gerstein, M. B. (2007). Toward a universal microarray: prediction of gene expression through nearest-neighbor probe sequence identification. *Nucleic Acids Res.* 35:e99. doi: 10.1093/nar/gkm549

Rubinov, M., Ypma, R. J. F., Watson, C., and Bullmore, E. T. (2015). Wiring cost and topological participation of the mouse brain connectome. *Proc. Natl. Acad. Sci. U.S.A.* 112, 10032–10037. doi: 10.1073/pnas.1420315112

Schröter, M., Paulsen, O., and Bullmore, E. T. (2017). Micro-connectomics: probing the organization of neuronal networks at the cellular scale. *Nat. Rev. Neurosci.* 18, 131–146. doi: 10.1038/nrn.2016.182

Schulze, A., and Downward, J. (2001). Navigating gene expression using microarrays–a technology review. *Nat. Cell Biol.* 3, E190–E195. doi: 10.1038/35087138

Seguin, C., van den Heuvel, M. P., and Zalesky, A. (2018). Navigation of brain networks. *Proc. Natl. Acad. Sci. U.S.A.* 115, 6297–6302 doi: 10.1073/pnas.1801351115

Seidlitz, J., Váša, F., Shinn, M., Romero-Garcia, R., Whitaker, K. J., Vértes, P. E., et al. (2018). Morphometric similarity networks detect microscale cortical organization and predict inter-individual cognitive variation. *Neuron* 97, 231–247.e7. doi: 10.1016/j.neuron.2017.11.039

Senden, M., Deco, G., de Reus, M. A., Goebel, R., and van den Heuvel, M. P. (2014). Rich club organization supports a diverse set of functional network configurations. *Neuroimage* 96, 174–182. doi: 10.1016/j.neuroimage.2014.03.066

Sinclair, B., Hansell, N. K., Blokland, G. A. M., Martin, N. G., Thompson, P. M., Breakspear, M., et al. (2015). Heritability of the network architecture of intrinsic brain functional connectivity. *Neuroimage* 121, 243–252. doi: 10.1016/j.neuroimage.2015.07.048

Smit, D. J. A., Stam, C. J., Posthuma, D., Boomsma, D. I., and de Geus, E. J. C. (2008). Heritability of "small-world" networks in the brain: a graph theoretical analysis of resting-state EEG functional connectivity. *Hum. Brain Mapp.* 29, 1368–1378. doi: 10.1002/hbm.20468

Song, H. F., Kennedy, H., and Wang, X.-J. (2014). Spatial embedding of structural similarity in the cerebral cortex. *Proc. Natl. Acad. Sci. U.S.A.* 111, 16580–16585. doi: 10.1073/pnas.1414153111

Sporns, O., Honey, C. J., and Kötter, R. (2007). Identification and classification of hubs in brain networks. *PLoS ONE* 2:e1049. doi: 10.1371/journal.pone.0001049

Sporns, O. (2011). The non-random brain: efficiency, economy, and complex dynamics. *Front. Comput. Neurosci.* 5:5. doi: 10.3389/fncom.2011.00005

Stephan, K. E., Kamper, L., Bozkurt, A., Burns, G. A., Young, M. P., and Kötter, R. (2001). Advanced database methodology for the Collation of Connectivity data on the Macaque brain (CoCoMac). *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 356, 1159–1186. doi: 10.1098/rstb.2001.0908

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.* 102, 15545–15550. doi: 10.1073/pnas.0506580102

Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., et al. (2009). mRNA-Seq whole-transcriptome analysis of a single cell. *Nat. Methods* 6, 377–382. doi: 10.1038/nmeth.1315

Tasic, B., Menon, V., Nguyen, T. N., Kim, T. K., Jarsky, T., Yao, Z., et al. (2016). Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nat. Neurosci.* 19, 335–346. doi: 10.1038/nn.4216

Tomasi, D., Wang, G. J., and Volkow, N. D. (2013). Energetic cost of brain functional connectivity. *Proc. Natl. Acad. Sci. U.S.A.* 110, 13642–13647. doi: 10.1073/pnas.1303346110

Towlson, E. K., Vértes, P. E., Ahnert, S. E., Schafer, W. R., and Bullmore, E. T. (2013). The rich club of the *C. elegans* neuronal connectome. *J. Neurosci.* 33, 6380–6387. doi: 10.1523/JNEUROSCI.3784-12.2013

Trabzuni, D., Ramasamy, A., Imran, S., Walker, R., Smith, C., Weale, M. E., et al. (2013). Widespread sex differences in gene expression and splicing in the adult human brain. *Nat. Commun.* 4:2771. doi: 10.1038/ncomms3771

Unger, E. R., Nitta, H., Lee, D. R., and Grogan, T. M. (2010). "*In situ* hybridization: principles and applications," in *Molecular Diagnostics*, eds W. W. Grody, R. M. Nakamura, C. M. Strom, and F. L. Kiechle (San Diego, CA: Elsevier), 71–79. doi: 10.1016/B978-0-12-369428-7.00007-0

Vaishnavi, S. N., Vlassenko, A. G., Rundle, M. M., Snyder, A. Z., Mintun, M. A., and Raichle, M. E. (2010). Regional aerobic glycolysis in the human brain. *Proc. Natl. Acad. Sci. U.S.A.* 107, 17757–17762. doi: 10.1073/pnas.1010459107

van den Heuvel, M. P., Bullmore, E. T., and Sporns, O. (2016a). Comparative connectomics. *Trends Cogn. Sci.* 20, 345–361. doi: 10.1016/j.tics.2016.03.001

van den Heuvel, M. P., Kahn, R. S., Goni, J., and Sporns, O. (2012). High-cost, high-capacity backbone for global brain communication. *Proc. Natl. Acad. Sci. U.S.A.* 109, 11372–11377. doi: 10.1073/pnas.1203593109

van den Heuvel, M. P., Scholtens, L. H., and de Reus, M. A. (2016b). Topological organization of connectivity strength in the rat connectome. *Brain Struct. Funct.* 221, 1719–1736. doi: 10.1007/s00429-015-0999-6

van den Heuvel, M. P., Sporns, O., Collin, G., Scheewe, T., Mandl, R. C., Cahn, W., et al. (2013b). Abnormal rich club organization and functional brain dynamics in schizophrenia. *JAMA Psychiatry* 70, 783–792. doi: 10.1001/jamapsychiatry.2013.1328

van den Heuvel, M. P., and Sporns, O. (2011). Rich-club organization of the human connectome. *J. Neurosci.* 31, 15775–15786. doi: 10.1523/JNEUROSCI.3539-11.2011

van den Heuvel, M. P., and Sporns, O. (2013). An anatomical substrate for integration among functional networks in human cortex. *J. Neurosci.* 33, 14489–14500. doi: 10.1523/JNEUROSCI.2128-13.2013

van den Heuvel, M. P., van Soelen, I. L., Stam, C. J., Kahn, R. S., Boomsma, D. I., and Hulshoff Pol, H. E. (2013a). Genetic control of functional brain network efficiency in children. *Eur. Neuropsychopharmacol.* 23, 19–23. doi: 10.1016/j.euroneuro.2012.06.007

Varadan, V., Miller, D. M., and Anastassiou, D. (2006). Computational inference of the molecular logic for synaptic connectivity in *C. elegans*. *Bioinformatics* 22, e497–e506. doi: 10.1093/bioinformatics/btl224

Varshney, L. R., Chen, B. L., Paniagua, E., Hall, D. H., and Chklovskii, D. B. (2011). Structural properties of the *Caenorhabditis elegans* neuronal network. *PLoS Comput. Biol.* 7:e1001066. doi: 10.1371/journal.pcbi.1001066

Vértes, P. E., Rittman, T., Whitaker, K. J., Romero-Garcia, R., Váša, F., Kitzbichler, M. G., et al. (2016). Gene transcription profiles associated with inter-modular hubs and connection distance in human functional magnetic resonance imaging networks. *Philos. Trans. R. Soc. B Biol. Sci.* 371, 735–769. doi: 10.1098/rstb.2015.0362

Vetere, G., Kenney, J. W., Tran, L. M., Xia, F., Steadman, P. E., Parkinson, J., et al. (2017). Chemogenetic interrogation of a brain-wide fear memory network in mice. *Neuron* 94, 363–374. doi: 10.1016/j.neuron.2017.03.037

Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* 10, 57–63. doi: 10.1038/nrg2484

Watts, D. J., and Strogatz, S. H. (1998). Collective dynamics of "small-world" networks. *Nature* 393, 440–442. doi: 10.1038/30918

Whitaker, K. J., Vértes, P. E., Romero-Garcia, R., Váša, F., Moutoussis, M., Prabhu, G., et al. (2016). Adolescence is associated with genomically patterned consolidation of the hubs of the human brain connectome. *Proc. Natl. Acad. Sci. U.S.A.* 113, 9105–9110. doi: 10.1073/pnas.1601745113

White, J. G., Southgate, E., Thomson, J. N., and Brenner, S. (1986). The structure of the nervous system of the nematode *Caenorhabditis elegans*. *Philos. Trans. R. Soc. B* 314, 1–340. doi: 10.1098/rstb.1986.0056

Wolf, L., Goldberg, C., Manor, N., Sharan, R., and Ruppin, E. (2011). Gene expression in the rodent brain is associated with its regional connectivity. *PLoS Comput. Biol.* 7:e1002040. doi: 10.1371/journal.pcbi.1002040

Yu, P., and Lin, W. (2016). Single-cell transcriptome study as big data. *Genom. Proteom. Bioinf.* 14, 21–30. doi: 10.1016/j.gpb.2016.01.005

Zalesky, A., Fornito, A., and Bullmore, E. (2012). On the use of correlation as a measure of network connectivity. *Neuroimage* 60, 2096–2106. doi: 10.1016/J.NEUROIMAGE.2012.02.001

Zamora-López, G., Zhou, C., and Kurths, J. (2010). Cortical hubs form a module for multisensory integration on top of the hierarchy of cortical networks. *Front. Neuroinform.* 4:1. doi: 10.3389/neuro.11.001.2010

Zhan, L., Jahanshad, N., Faskowitz, J., Zhu, D., Prasad, G., Martin, N. G., et al. (2015). Heritability of brain network topology in 853 twins and siblings. *Proc. IEEE Int. Symp. Biomed. Imaging* 2015, 449–453 doi: 10.1109/ISBI.2015.7163908

Zhong, S., Zhang, S., Fan, X., Wu, Q., Yan, L., Dong, J., et al. (2018). A single-cell RNA-seq survey of the developmental landscape of the human prefrontal cortex. *Nature* 555, 524–528. doi: 10.1038/nature25980

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: info@frontiersin.org | +41 21 510 17 00

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership